Lecture 17:
Robot Learning

So far: Supervised Learning

Supervised Learning

Data: (x, y)

x is data, y is label

Goal: Learn a *function* to map x -> y

Examples: Classification, regression, object detection, semantic segmentation, image captioning, etc.

Classification

Cat

So far: Self-Supervised Learning

Self-Supervised Learning

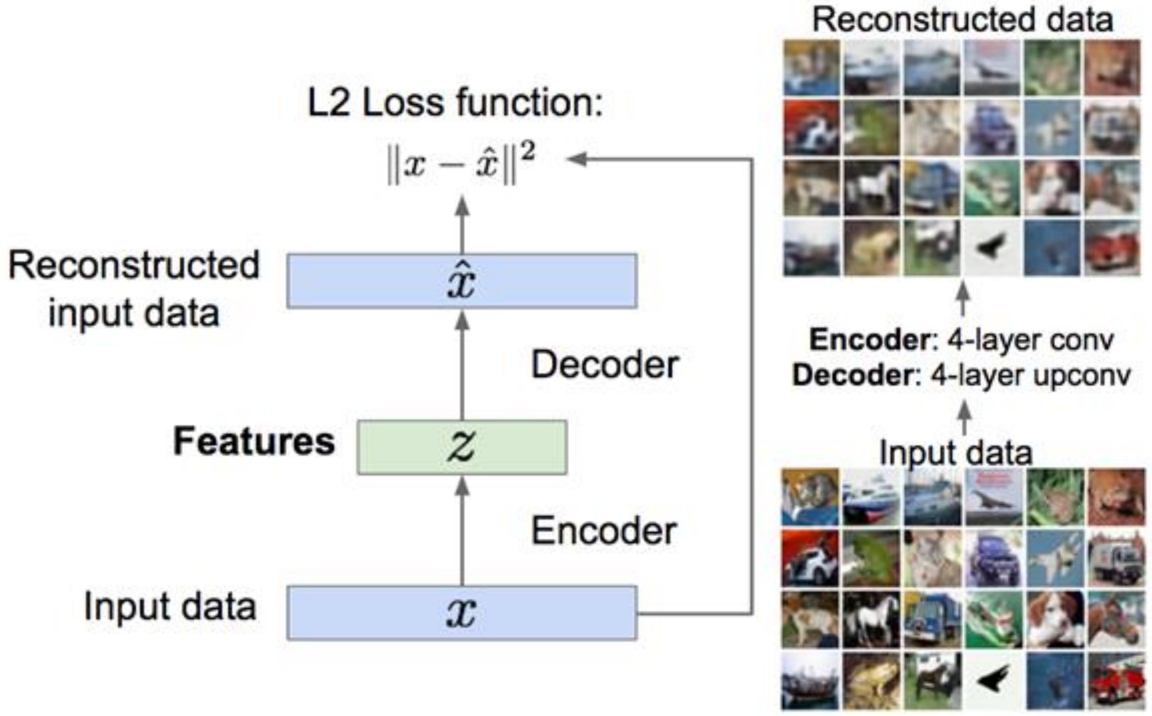
Data: x

Just data, no labels!

Goal: Learn some underlying hidden *structure* of the data

Examples: Clustering, dimensionality reduction, feature learning, density estimation, etc.

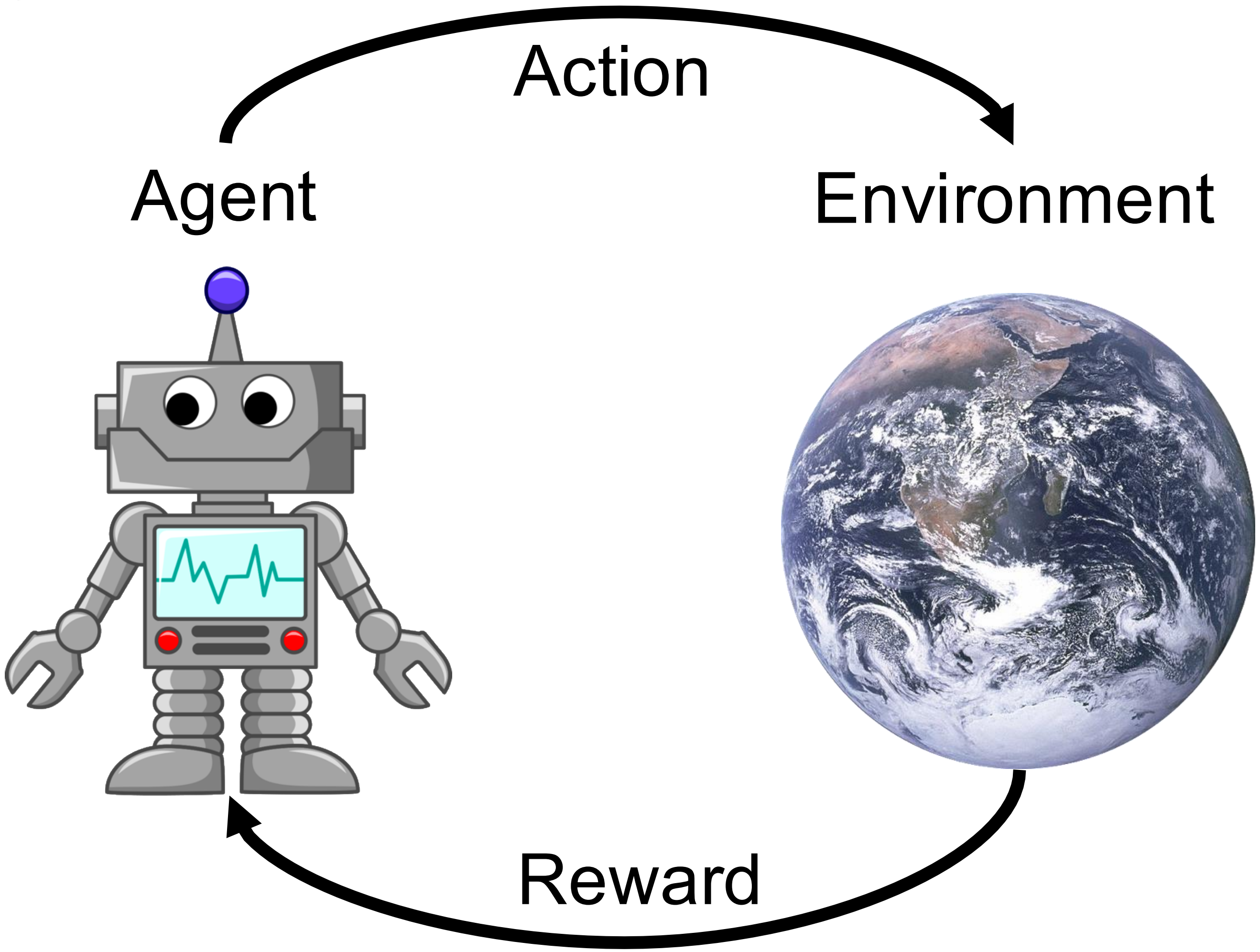
Feature Learning
(e.g., autoencoders)

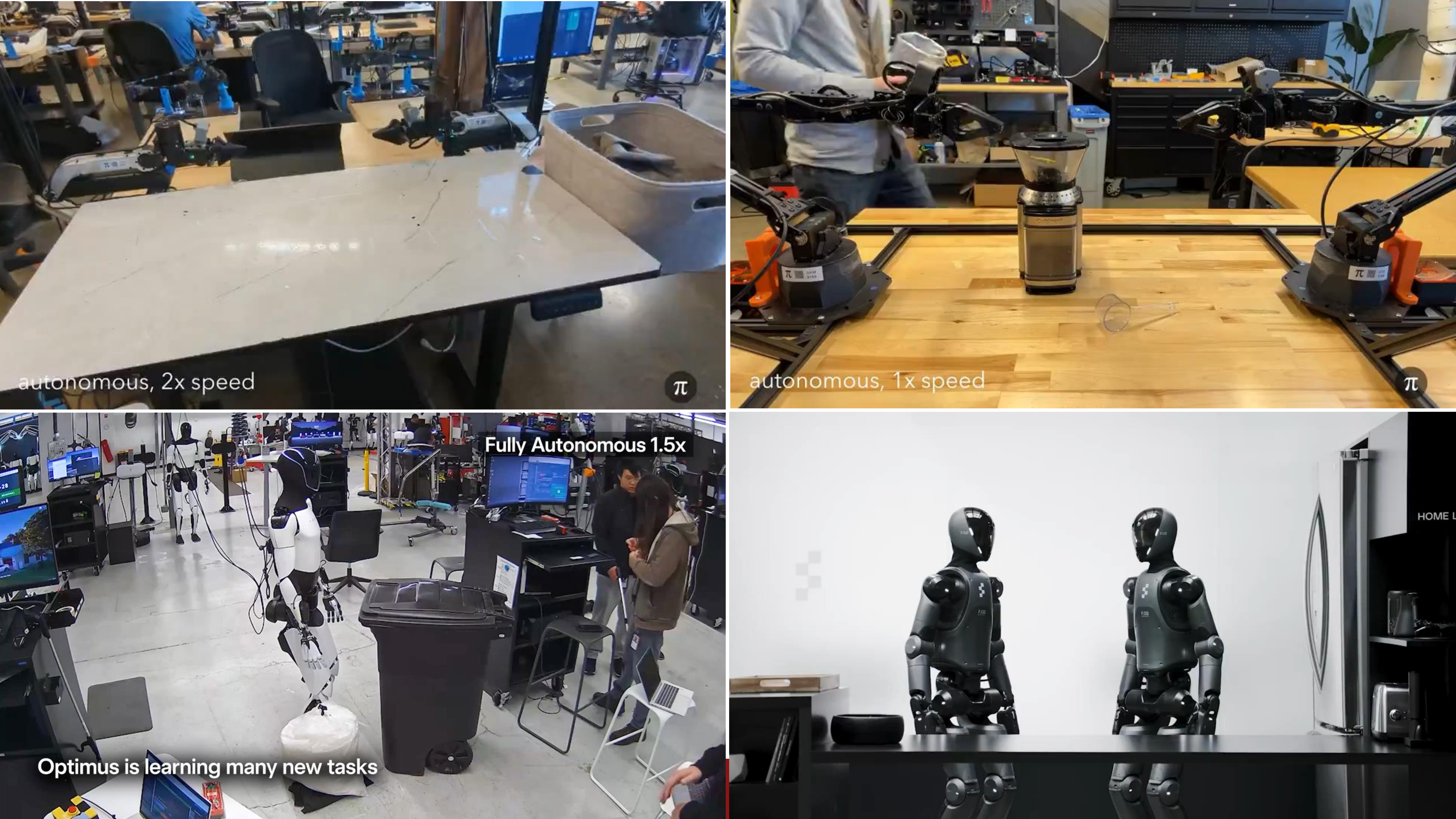


Today: Robot Learning

Problems where an **agent** performs **actions** in the **environment**, and receives **rewards**

Goal: Learn how to take actions that maximize reward




autonomous, 2x speed
autonomous, 1x speed
Fully Autonomous 1.5x
Optimus is learning many new tasks

A Fast-Growing Field

Reuters World Business Markets Sustainability More

Robot AI startup Physical Intelligence raises $400 mln from Bezos, OpenAI

By Reuters

November 4, 2024 12:38 PM EST · Updated 3 months ago

Aa

Series B: 1X Secures $100M Funding

January 11, 2024 Author: 1X

Skild AI grabs $300M to build foundation model for robotics

By Mike Oitzman | July 10, 2024

From self-driving cars to chore-battling bots: Robot Guru Kyle Vogt raises $150M for The Bot Company

BY VIVEK CHHETRI · MAY 14, 2024 · 2 MINUTE READ

Reuters World Business Markets Sustainability More

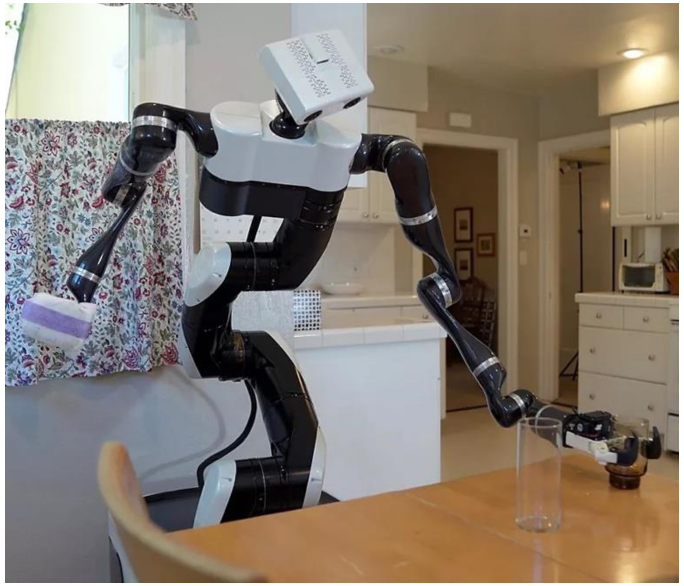
Robotics startup Figure raises $675 mln from Microsoft, Nvidia, OpenAI

By Harshita Mary Varghese and Krystal Hu

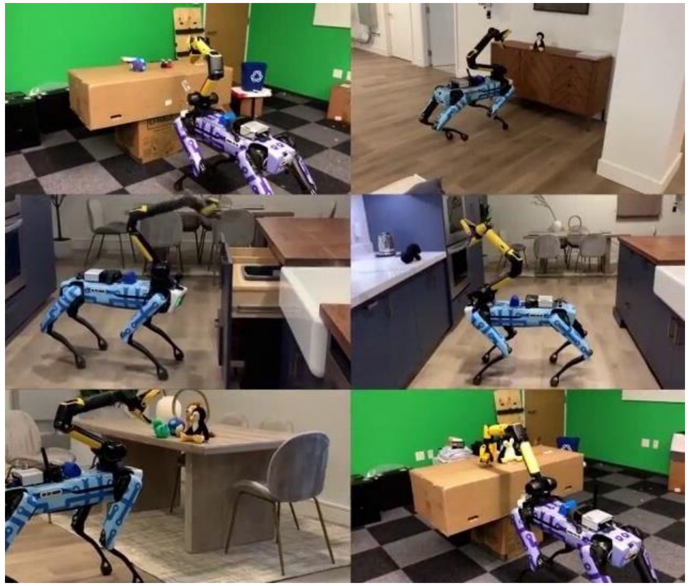
February 29, 2024 11:20 AM EST · Updated a year ago

Aa

A Fast-Growing Field



Toyota Research Institute



Meta AI Research

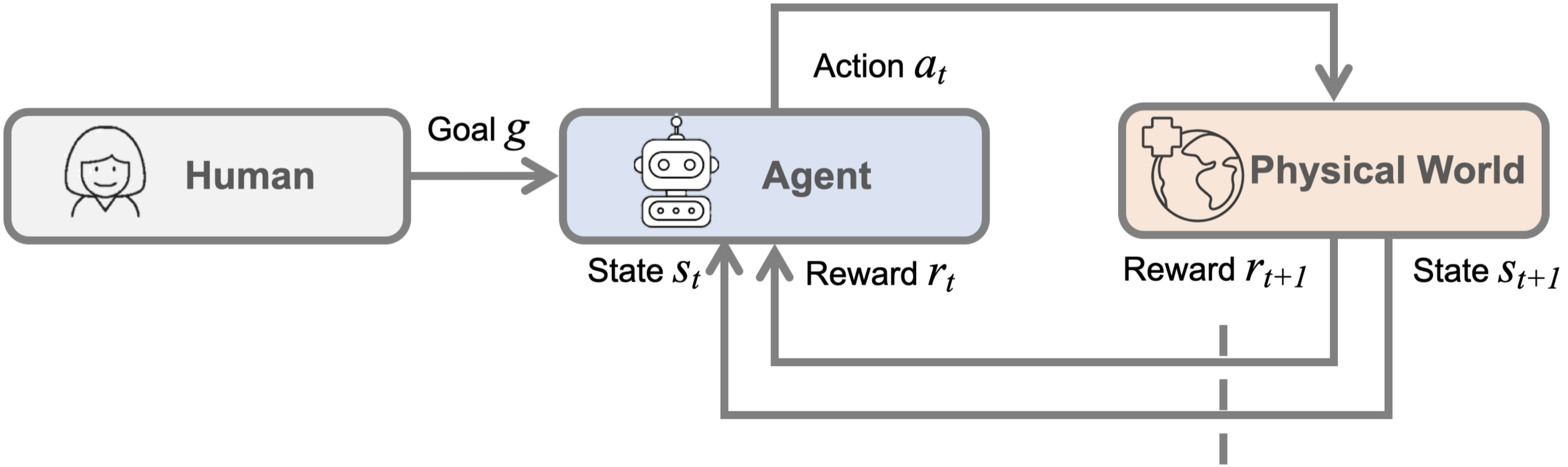
Google Robotics

Nvidia Research

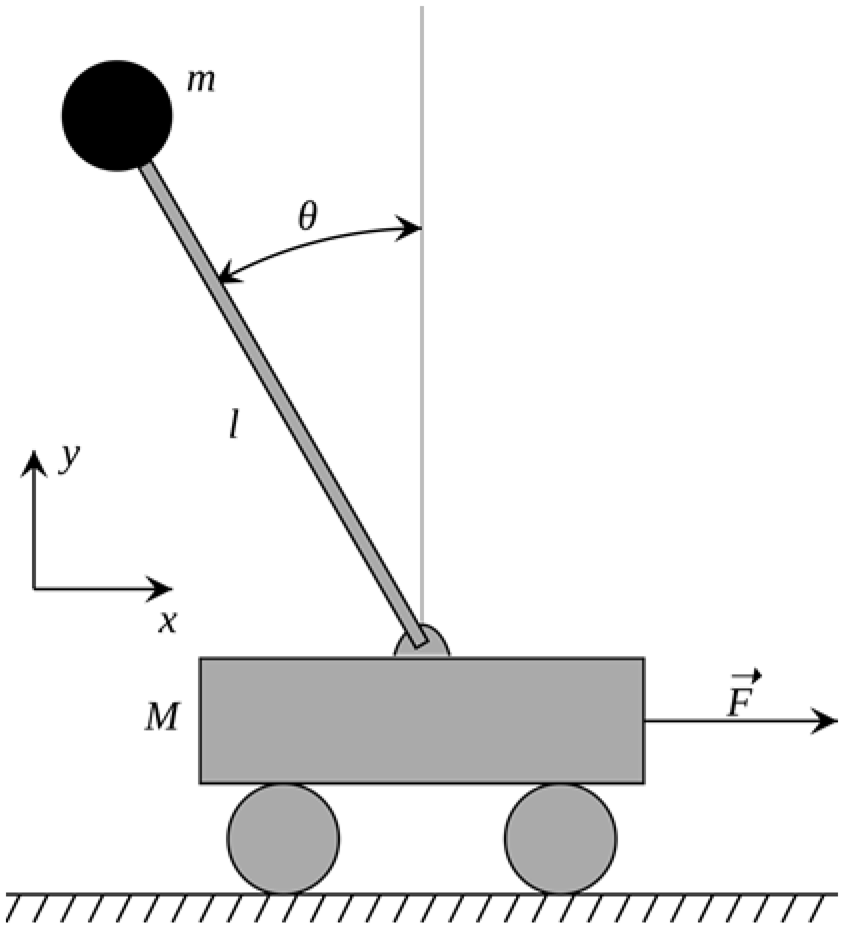
Overview

- Problem formulation
- Robot perception
- Reinforcement learning
- Model learning & model-based planning
- Imitation learning
- Robotic foundation models
- Remaining challenges

Problem Formulation



Example: Cart-Pole Problem



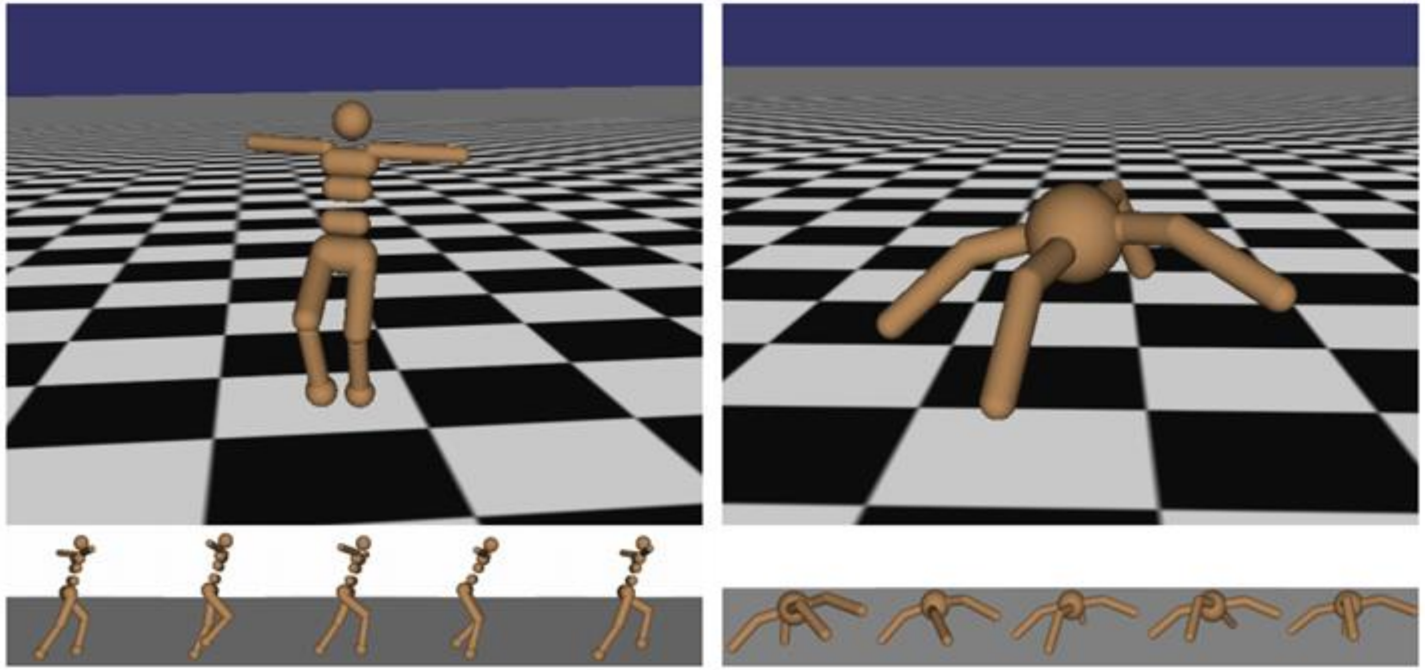
Goal: Balance a pole on top of a movable cart

State: angle, angular speed, position, horizontal velocity

Action: horizontal force applied to the cart

Reward: 1 at each time step if the pole is upright

Example: Robot Locomotion



Goal: Make the robot move forward

State: Angle, position, velocity of all joints

Action: Torques applied to joints

Reward: 1 at each time step upright + forward movement

Figure from: Schulman et al, "High-Dimensional Continuous Control Using Generalized Advantage Estimation", ICLR 2016

Example: Atari Games

Goal: Complete the game with the highest score

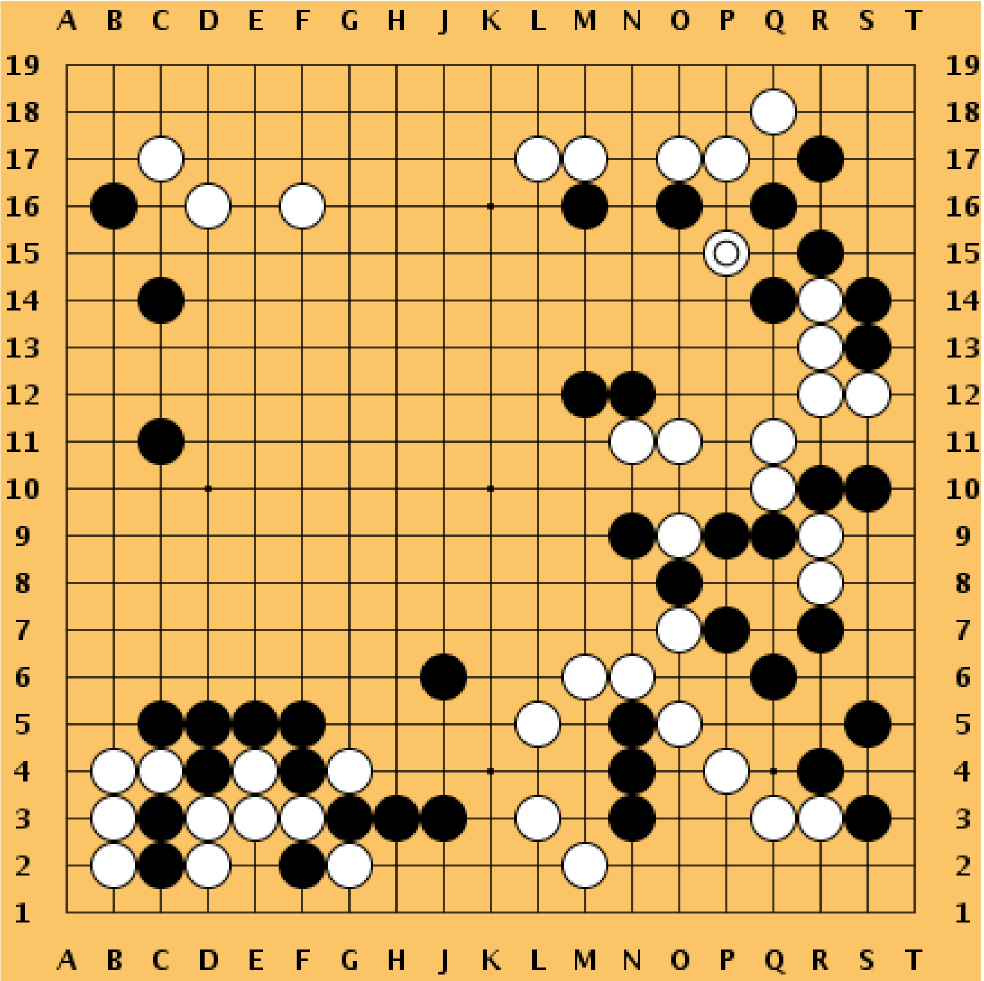
State: Raw pixel inputs of the game screen

Action: Game controls e.g. Left, Right, Up, Down

Reward: Score increase/decrease at each time step

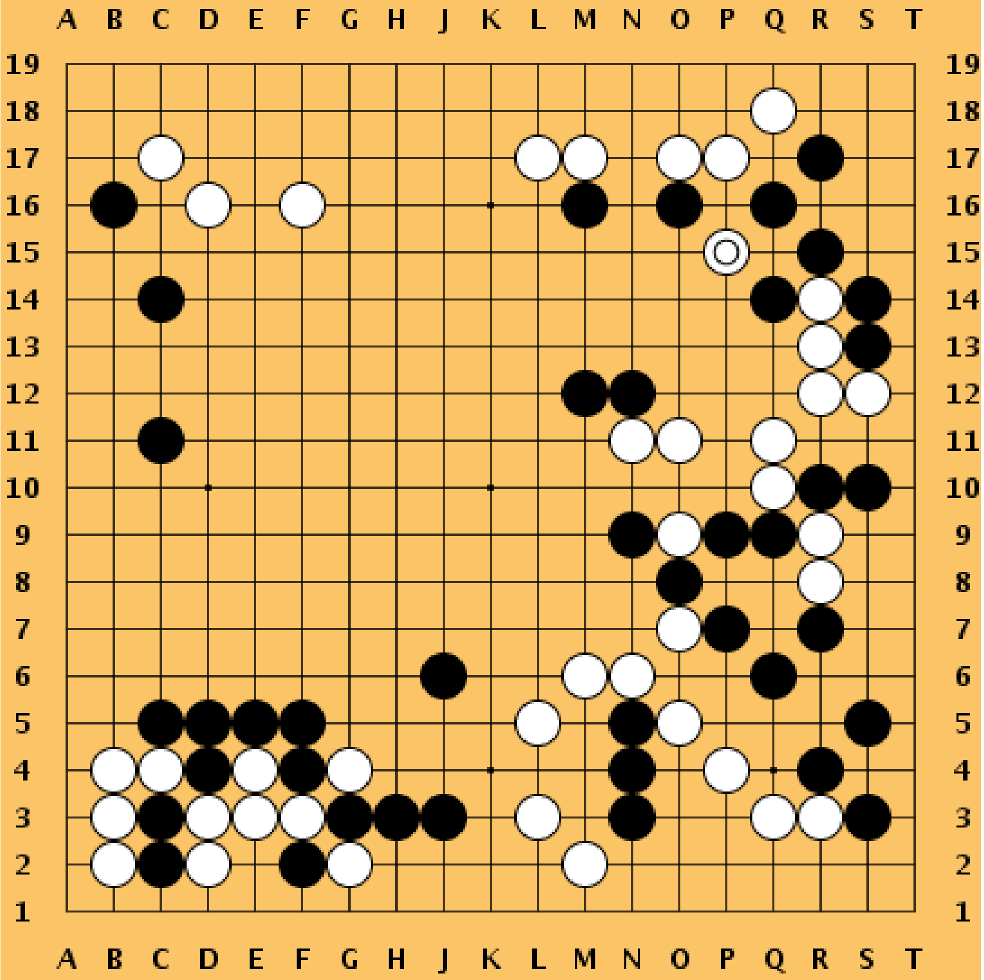
Mnih et al, “Playing Atari with Deep Reinforcement Learning”, NeurIPS Deep Learning Workshop, 2013

Example: Go



Goal: Win the game!

Example: Go



Goal: Win the game!

State: Position of all pieces

Action: Where to put the next piece down

Reward: On last turn: 1 if you won, 0 if you lost

Example: Text Generation

Goal: Predict the next word!

<s> CS231n
midterm
was ___

Example: Text Generation

<s> CS231n
midterm
was ___

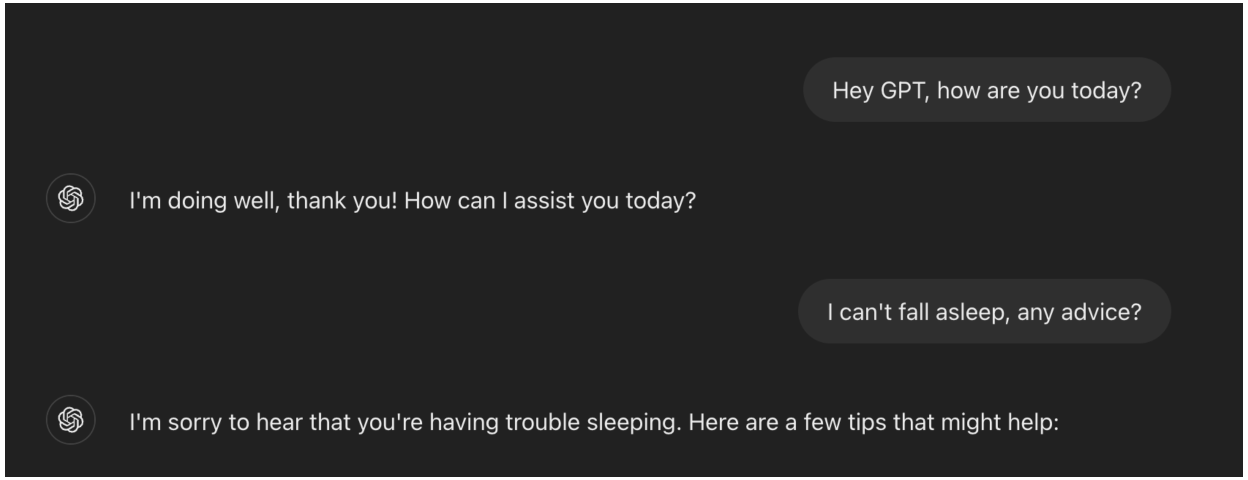
Goal: Predict the next word!

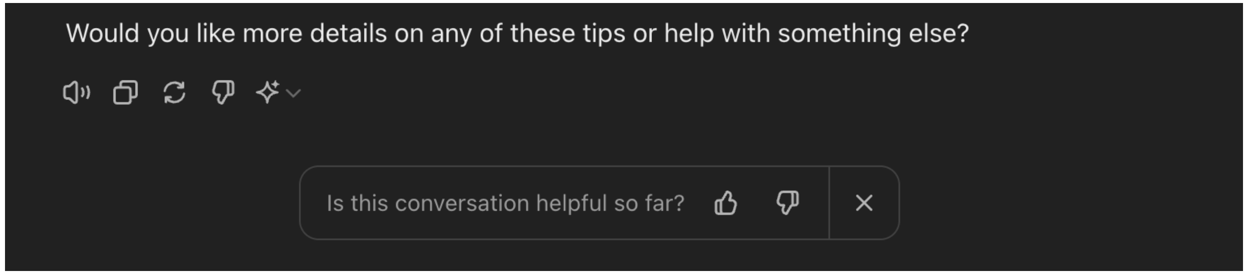
State: Current words in the sentence

Action: Next word

Reward: 1 if correct, 0 otherwise

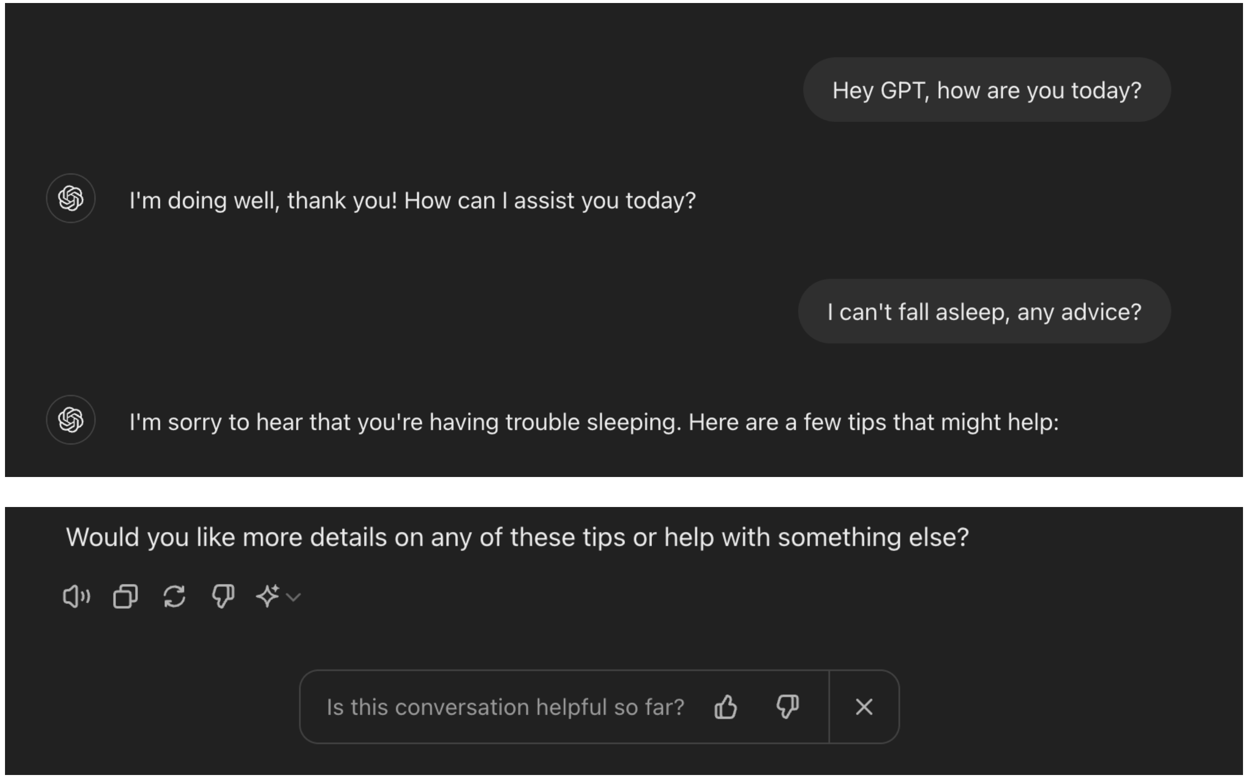
Example: Chatbot





Goal: Be a good companion!

Example: Chatbot



Goal: Be a good companion!

State: Current conversation

Action: Next sentence

Reward: Human evaluation, 1 if satisfied, -1 if unsatisfied, 0 neutral

Example: Cloth folding robot

Goal: Fold the cloth

Example: Cloth folding robot

Goal: Fold the cloth

State: Current conversation

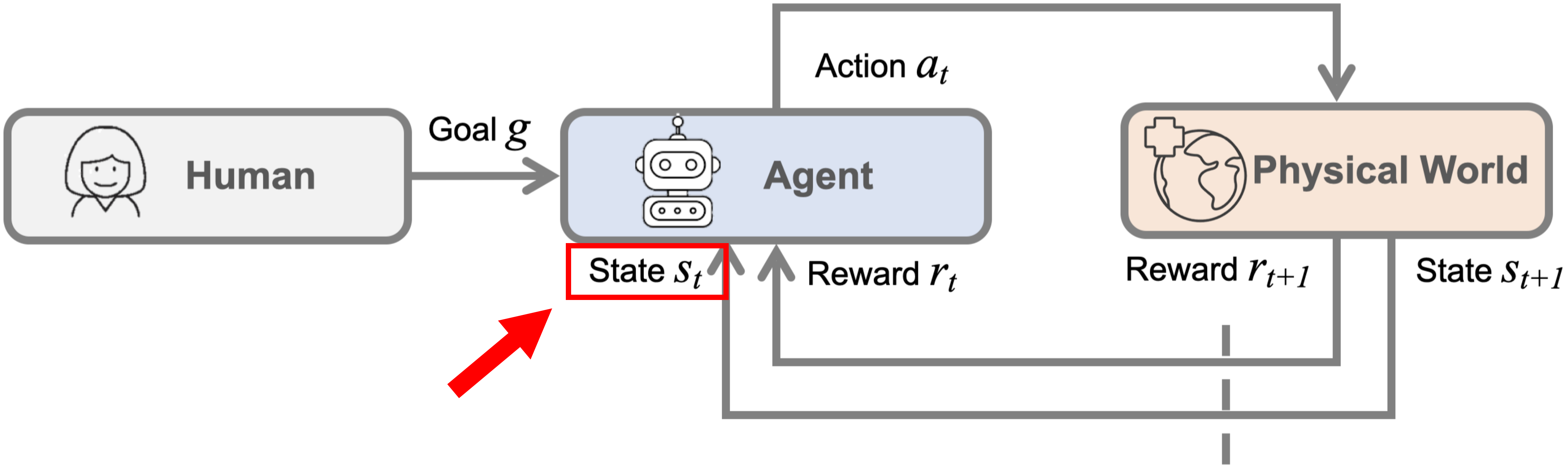
Action: Robot end-effector motions

Reward: Human evaluation, 1 if cloth is folded, 0 otherwise

Overview

- Problem formulation
- Robot perception
- Reinforcement learning
- Model learning & model-based planning
- Imitation learning
- Robotic foundation models
- Remaining challenges

What is Robot Perception?



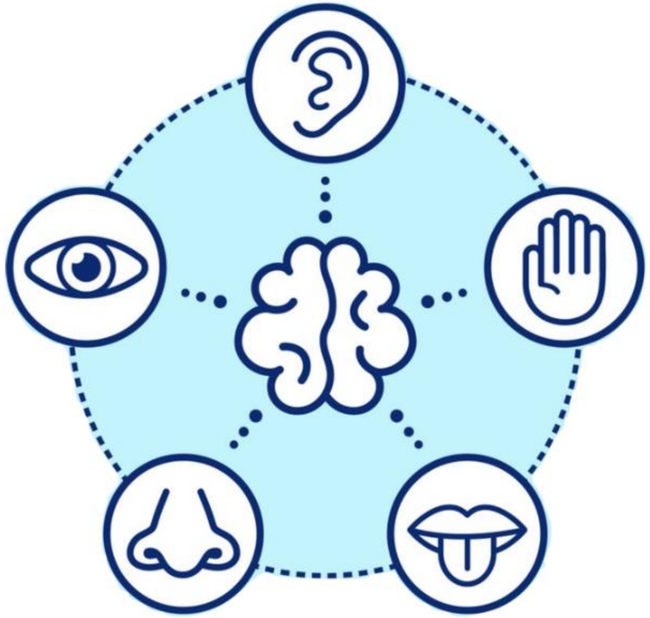
What is Robot Perception?

Making sense of the unstructured real world …

- Incomplete knowledge of objects and scenes
- Imperfect actions may lead to failure
- Environment dynamics and other agents

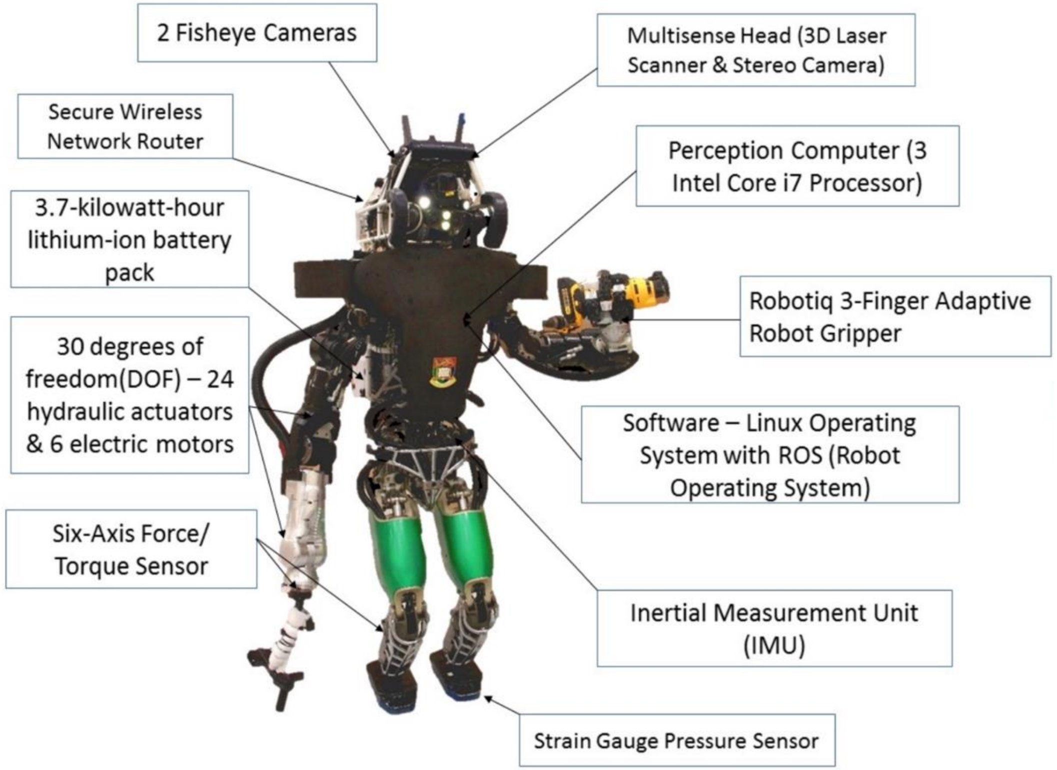
Sensors for Robotics

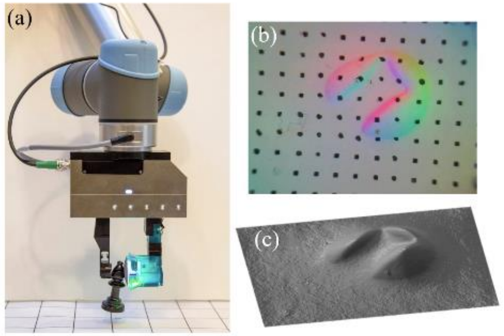
Understanding the interactions with the world through multimodal senses



Sensors for Robotics

Understanding the interactions with the world through multimodal senses

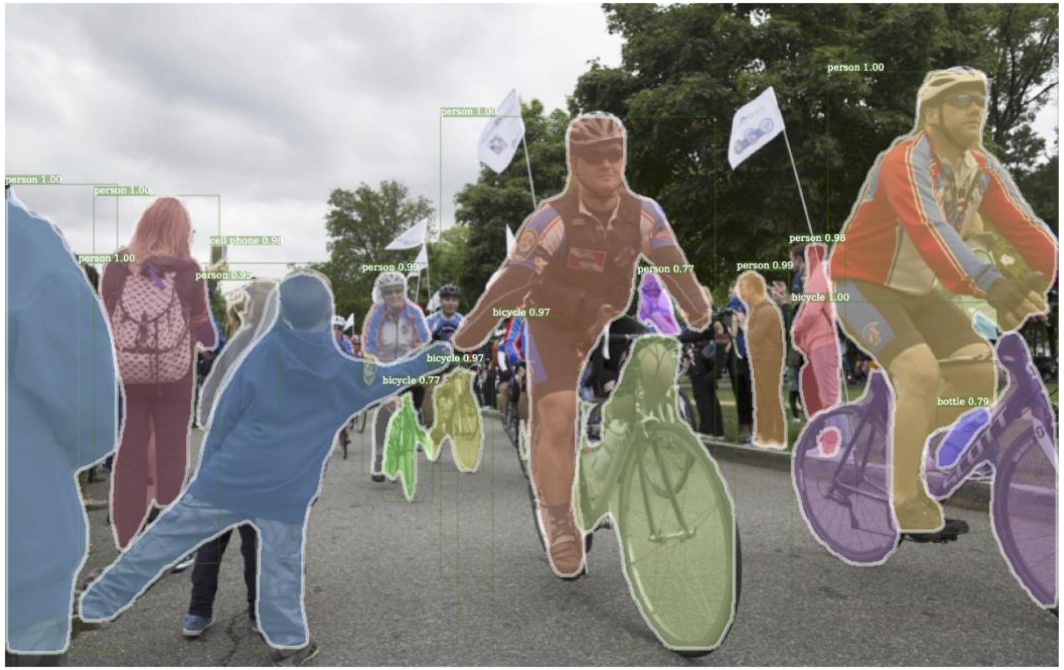




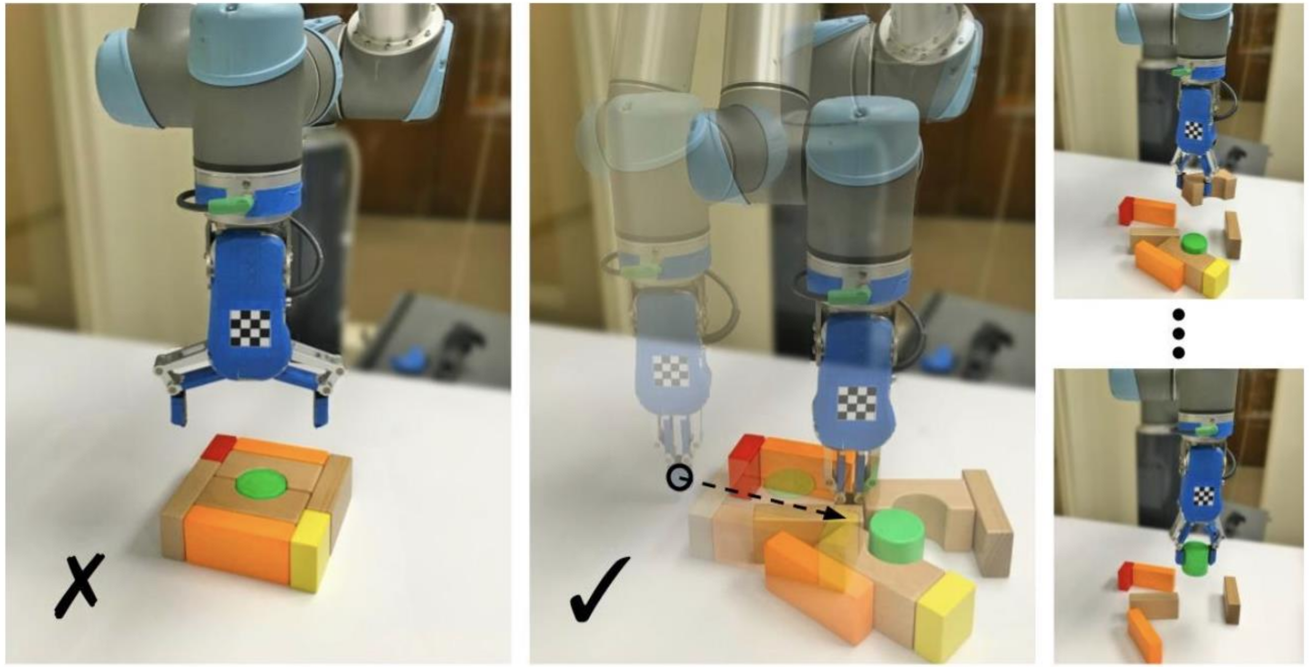
[Source: HKU Advanced Robotics Laboratory]

Robot Vision vs. Computer Vision

Robot vision is embodied, active, and environmentally situated.



[Detectron - Facebook AI Research]



[Zeng et al., IROS 2018]

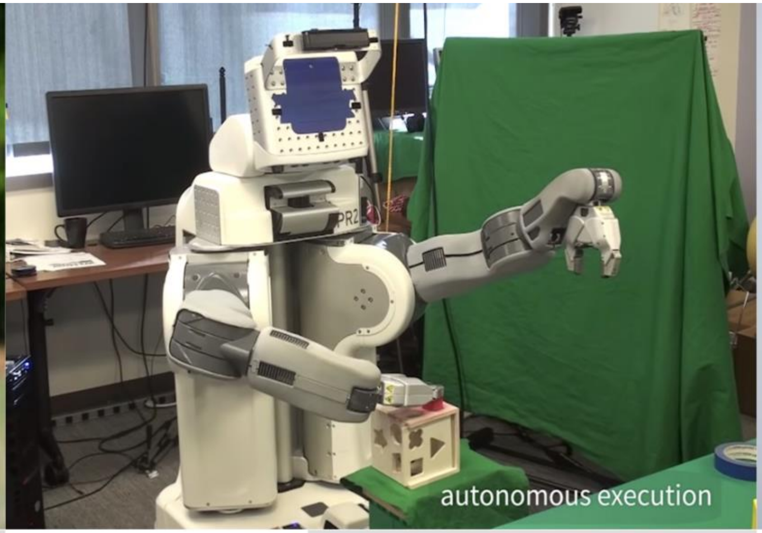
Robot Vision vs. Computer Vision

Robot vision is embodied, active, and environmentally situated.

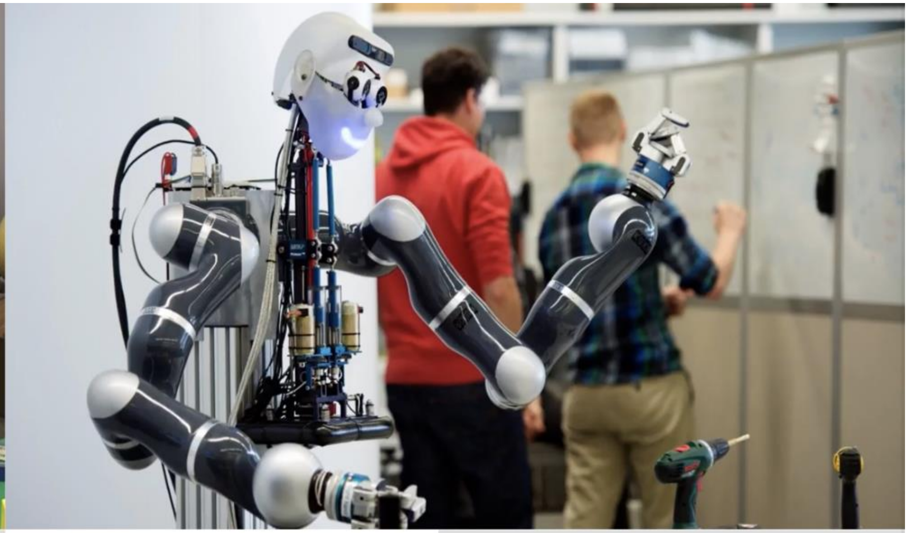
- **Embodied:** Robots have physical bodies and experience the world directly. Their actions are part of a dynamic with the world and have immediate feedback on their own sensation.
- **Active:** Robots are active perceivers. It knows why it wishes to sense, and chooses what to perceive, and determines how, when and where to achieve that perception.
- **Situated:** Robots are situated in the world. They do not deal with abstract descriptions, but with the "here" and "now" of the world directly influencing the behavior of the system.

The Perception-Action Loop

[Sa et al. IROS 2014]

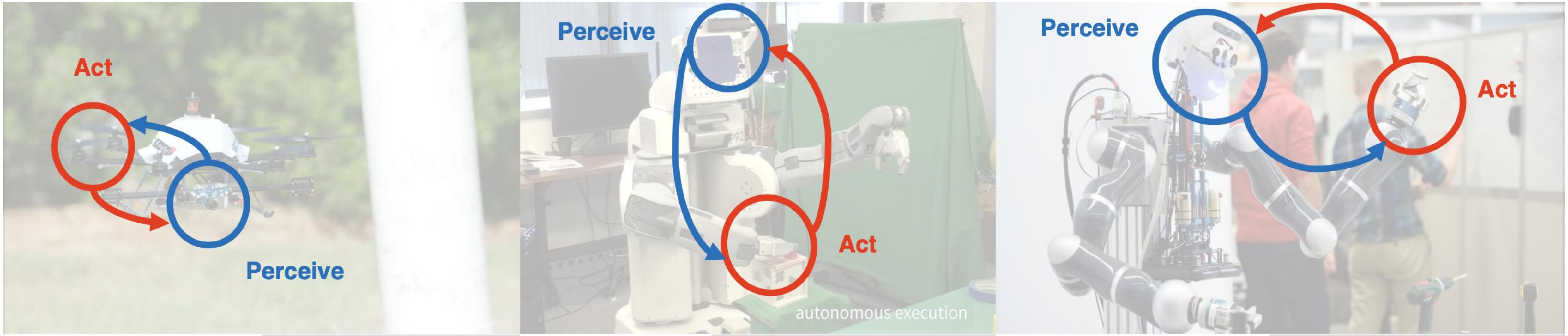


[Levine et al. JMLR 2016]



[Bohg et al. ICRA 2018]

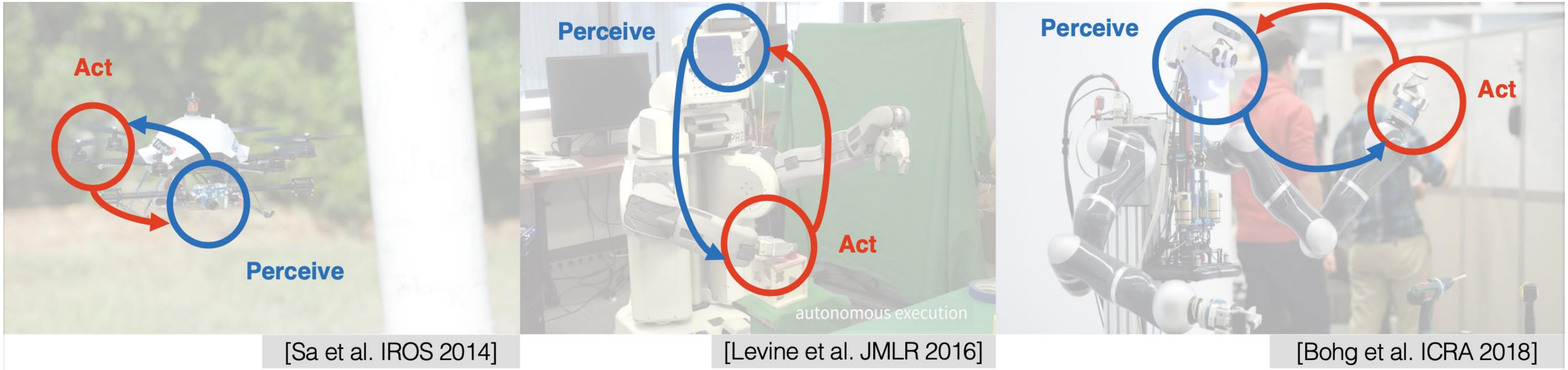
The Perception-Action Loop



[Sa et al. IROS 2014] [Levine et al. JMLR 2016] [Bohg et al. ICRA 2018]

The Perception-Action Loop

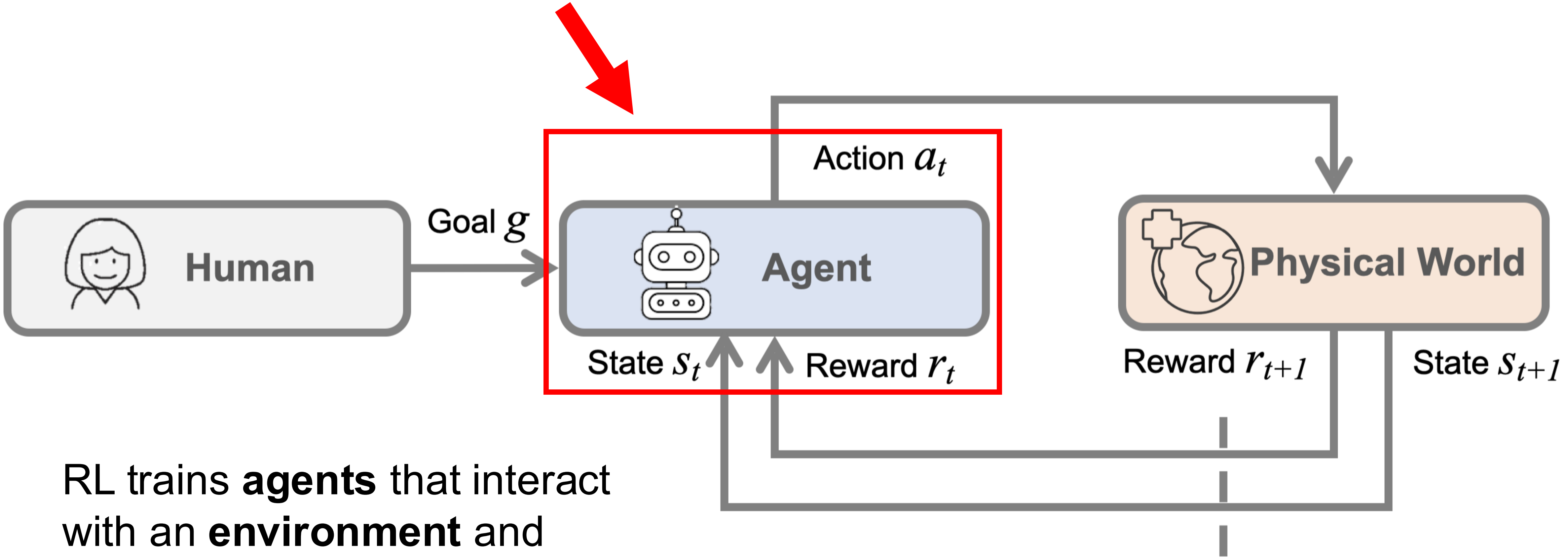
A key challenge in Robot Learning is to close the perception-action loop.



Overview

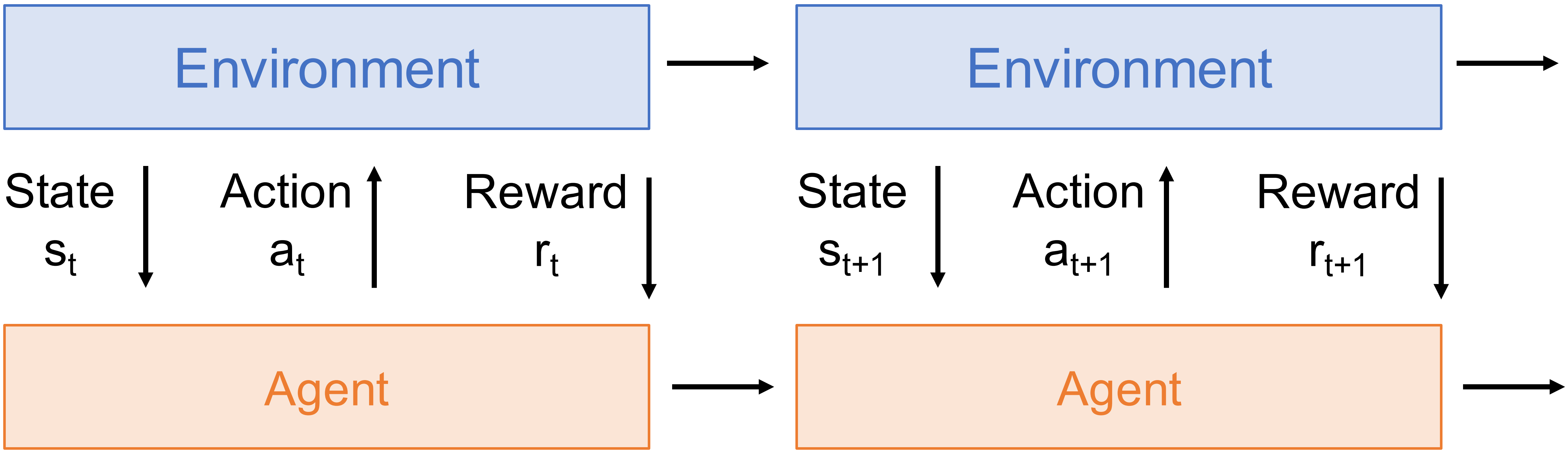
- Problem formulation
- Robot perception
- Reinforcement learning
- Model learning & model-based planning
- Imitation learning
- Robotic foundation models
- Remaining challenges

Reinforcement Learning

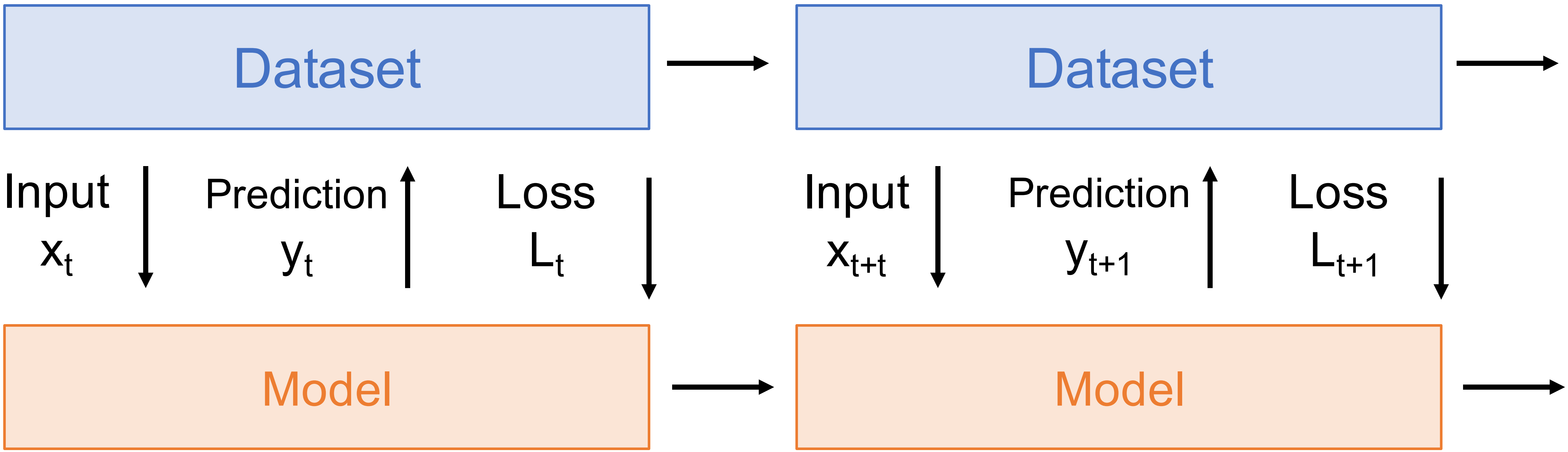


RL trains **agents** that interact with an **environment** and learn to maximize **reward** **(trial and error)**

Reinforcement Learning vs Supervised Learning

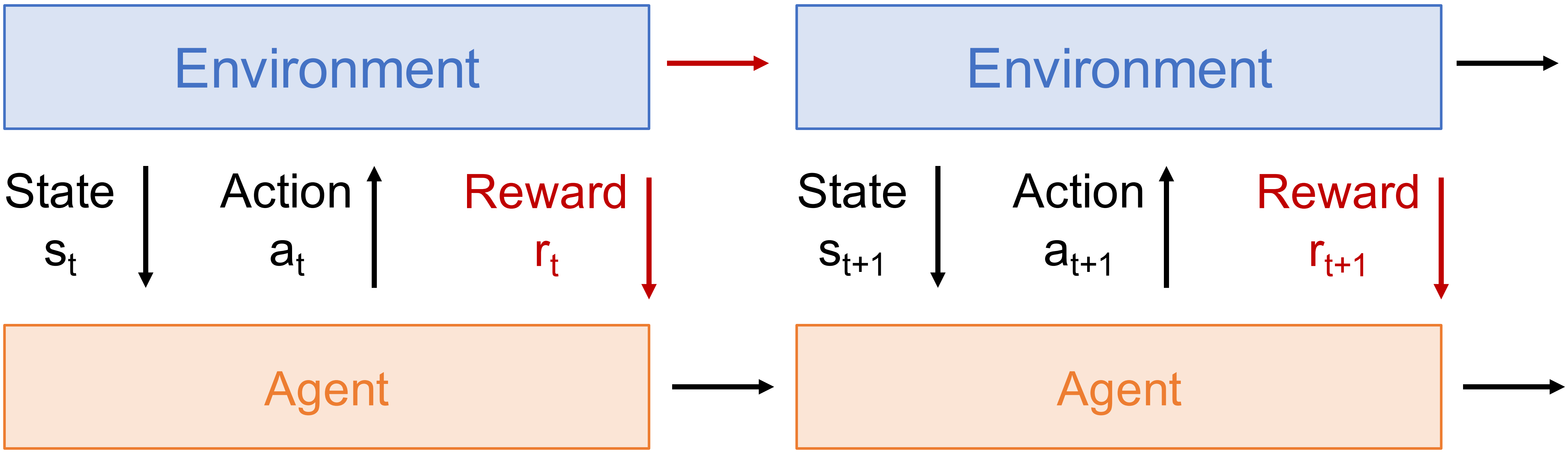


Reinforcement Learning vs Supervised Learning



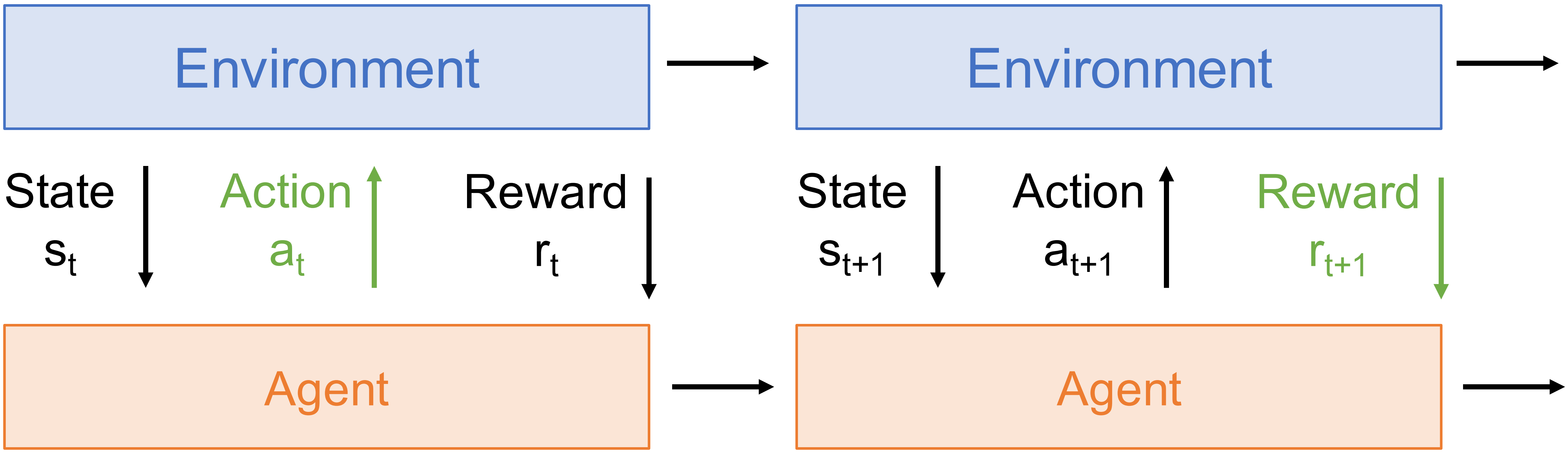
Why is RL different from normal supervised learning?

Reinforcement Learning vs Supervised Learning



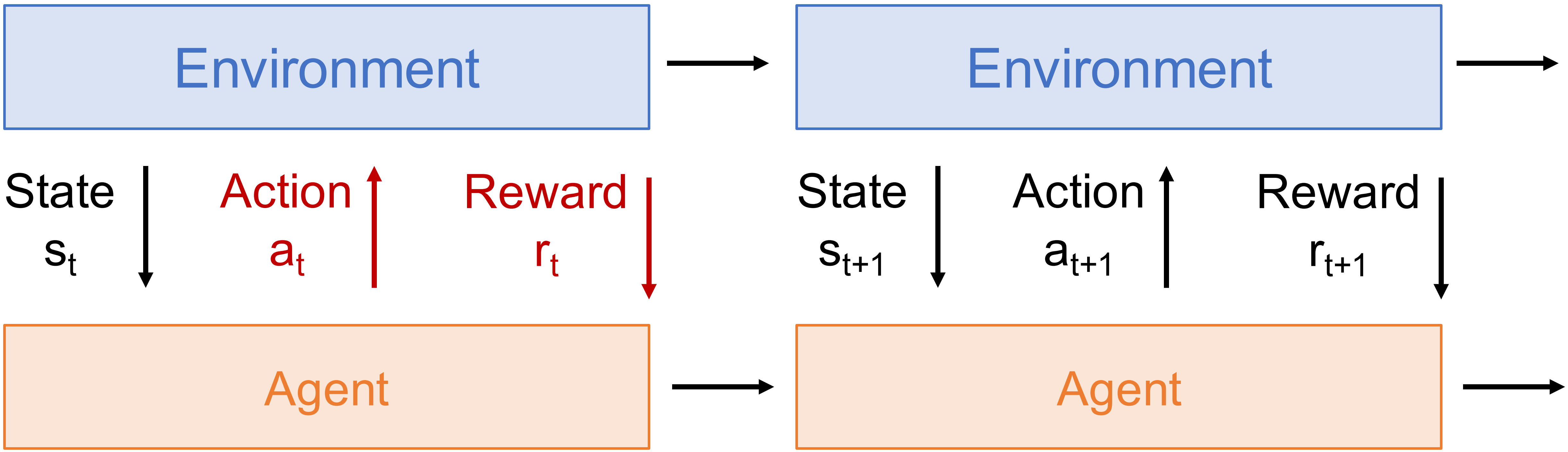
Stochasticity: Rewards and state transitions may be random

Reinforcement Learning vs Supervised Learning



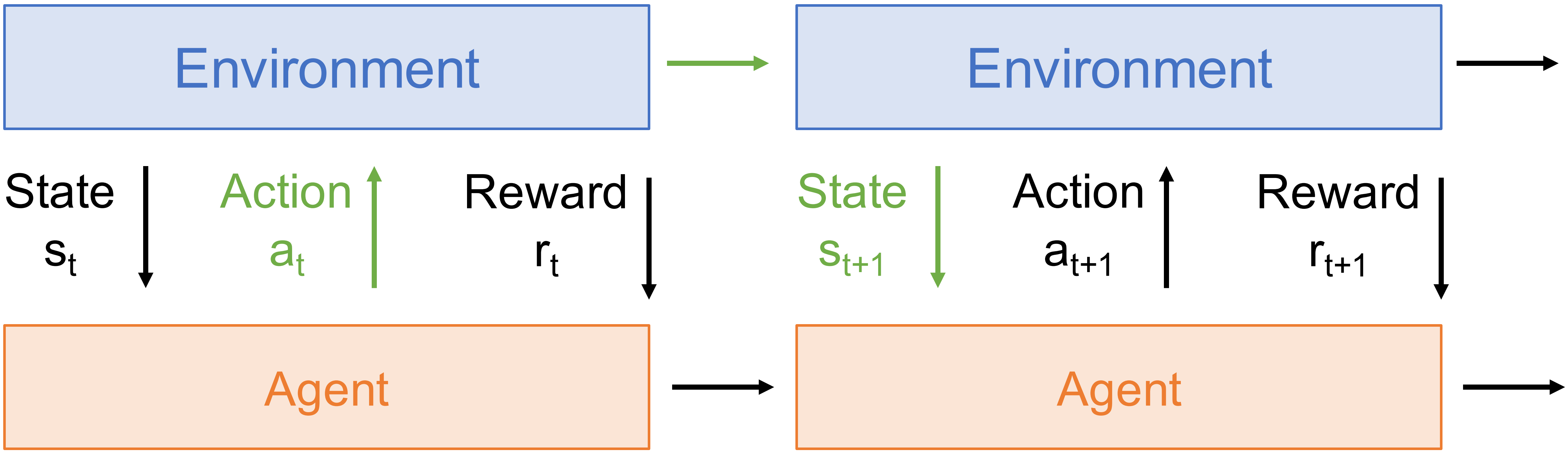
Credit assignment: Reward r_t may not directly depend on action a_t

Reinforcement Learning vs Supervised Learning



Nondifferentiable: Can't backprop through world; can't compute dr_t/da_t

Reinforcement Learning vs Supervised Learning



Nonstationary: What the agent experiences depends on how it acts

Case Study: Playing Atari Games

Goal: Complete the game with the highest score

State: Raw pixel inputs of the game screen
Action: Game controls e.g. Left, Right, Up, Down
Reward: Score increase/decrease at each time step

Mnih et al, “Playing Atari with Deep Reinforcement Learning”, NeurIPS Deep Learning Workshop, 2013

Case Study: Playing Atari Games

Mnih et al, “Playing Atari with Deep Reinforcement Learning”, NeurIPS Deep Learning Workshop, 2013

$Q(s, a; \theta)$
Neural network with weights θ

Network output:
Q-values for all actions

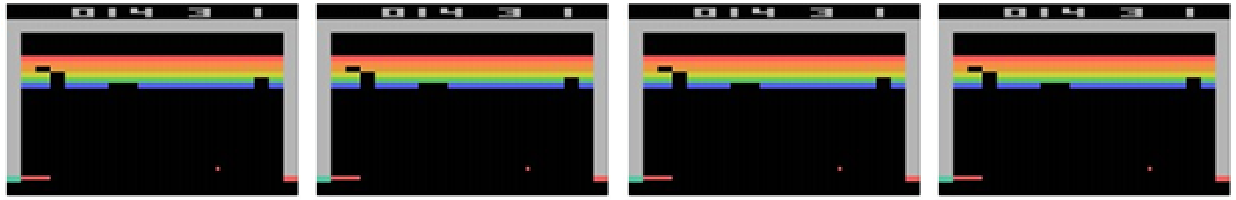
With 4 actions: last layer gives values $Q(s_t, a_1)$, $Q(s_t, a_2)$, $Q(s_t, a_3)$, $Q(s_t, a_4)$

FC-A (Q-values)

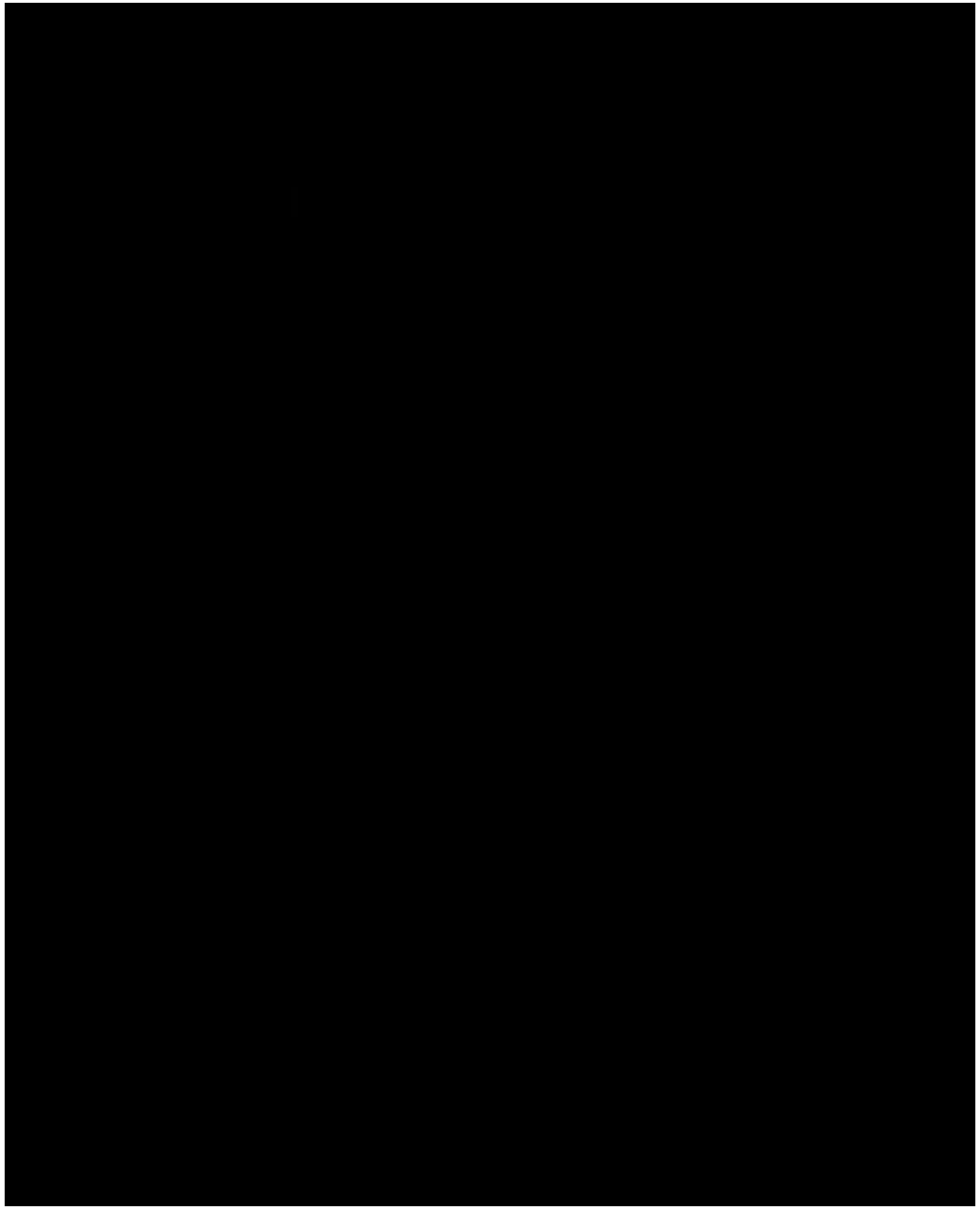
FC-256

Conv(16->32, 4x4, stride 2)

Conv(4->16, 8x8, stride 4)



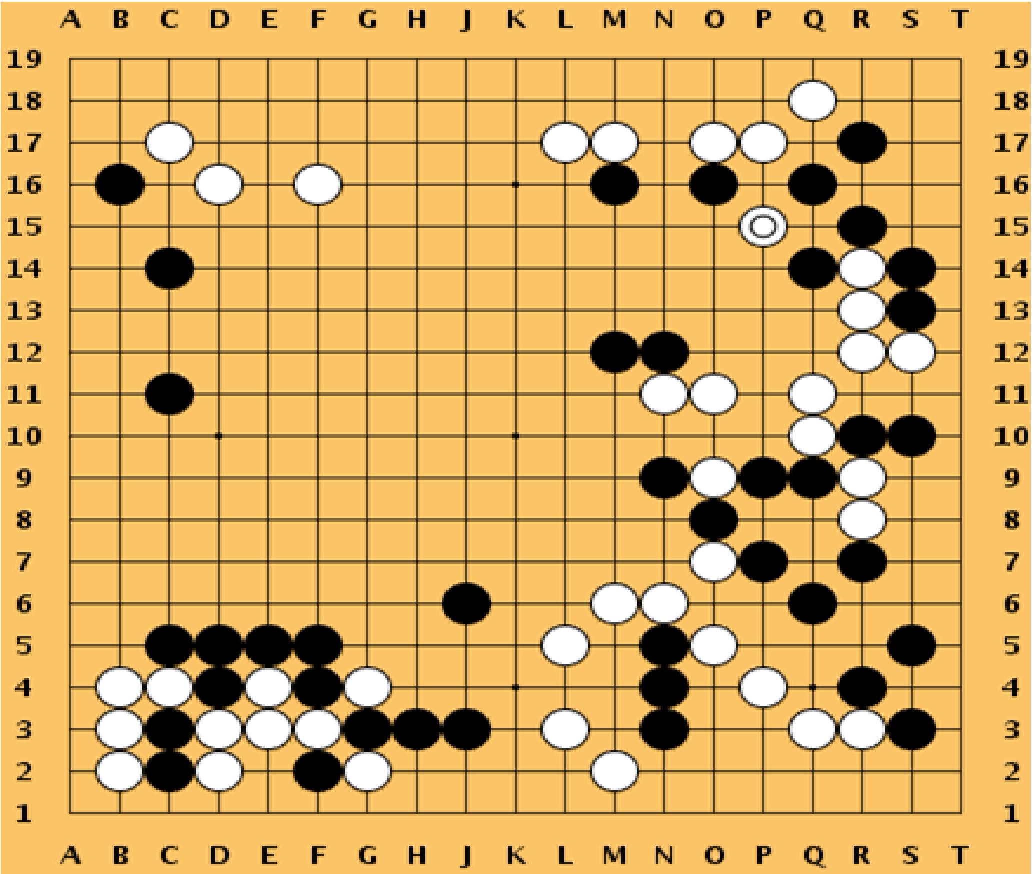
Network input: state s_t: 4x84x84 stack of last 4 frames
(after RGB->grayscale conversion, downsampling, and cropping)



https://www.youtube.com/watch?v=V1eYniJ0Rnk

Case Study: Playing Games

AlphaGo: (January 2016)
- Used imitation learning + tree search + RL
- Beat 18-time world champion Lee Sedol

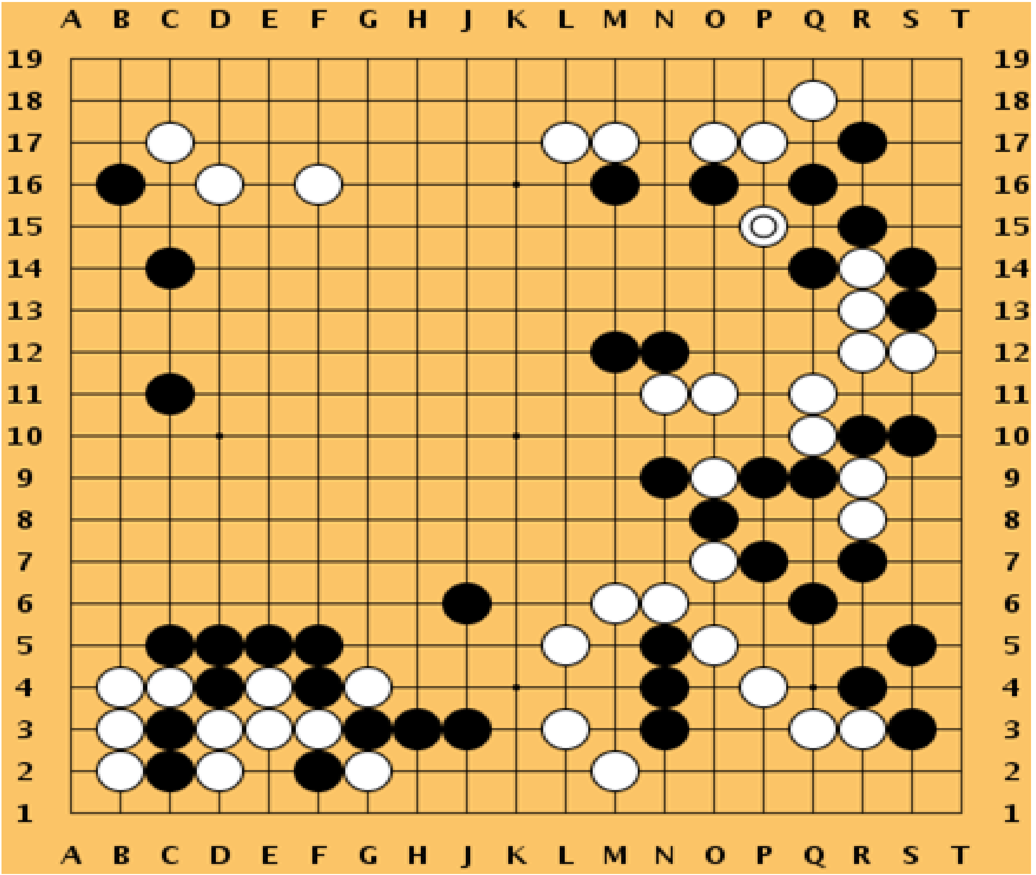


Silver et al, "Mastering the game of Go with deep neural networks and tree search", Nature 2016
Silver et al, "Mastering the game of Go without human knowledge", Nature 2017
Silver et al, "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play", Science 2018
Schrittwieser et al, "Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model", arXiv 2019

Case Study: Playing Games

AlphaGo: (January 2016)
- Used imitation learning + tree search + RL
- Beat 18-time world champion Lee Sedol

AlphaGo Zero (October 2017)
- Simplified version of AlphaGo
- No longer using imitation learning
- Beat (at the time) #1 ranked Ke Jie



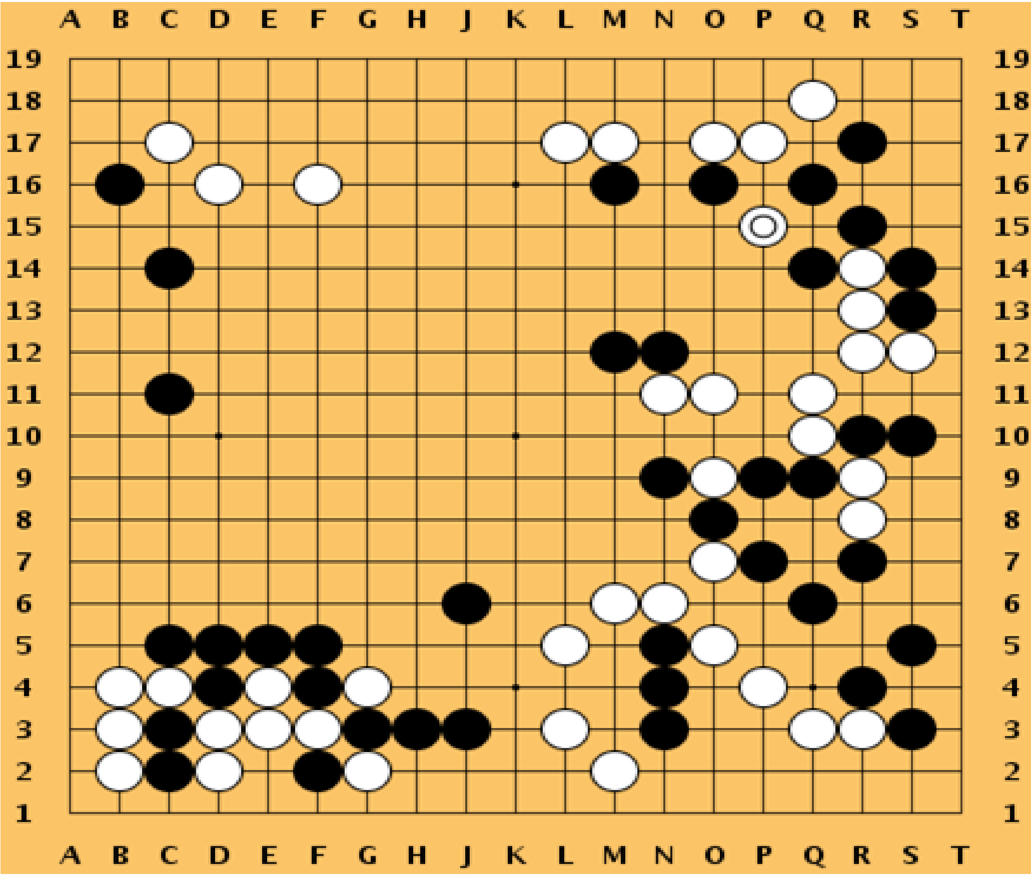
Silver et al, “Mastering the game of Go with deep neural networks and tree search”, Nature 2016
Silver et al, “Mastering the game of Go without human knowledge”, Nature 2017
Silver et al, “A general reinforcement learning algorithm that masters chess, shogi, and go through self-play”, Science 2018
Schrittwieser et al, “Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model”, arXiv 2019

Case Study: Playing Games

AlphaGo: (January 2016)
- Used imitation learning + tree search + RL
- Beat 18-time world champion Lee Sedol

AlphaGo Zero (October 2017)
- Simplified version of AlphaGo
- No longer using imitation learning
- Beat (at the time) #1 ranked Ke Jie

Alpha Zero (December 2018)
- Generalized to other games: Chess and Shogi



Silver et al, “Mastering the game of Go with deep neural networks and tree search”, Nature 2016
Silver et al, “Mastering the game of Go without human knowledge”, Nature 2017
Silver et al, “A general reinforcement learning algorithm that masters chess, shogi, and go through self-play”, Science 2018
Schrittwieser et al, “Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model”, arXiv 2019

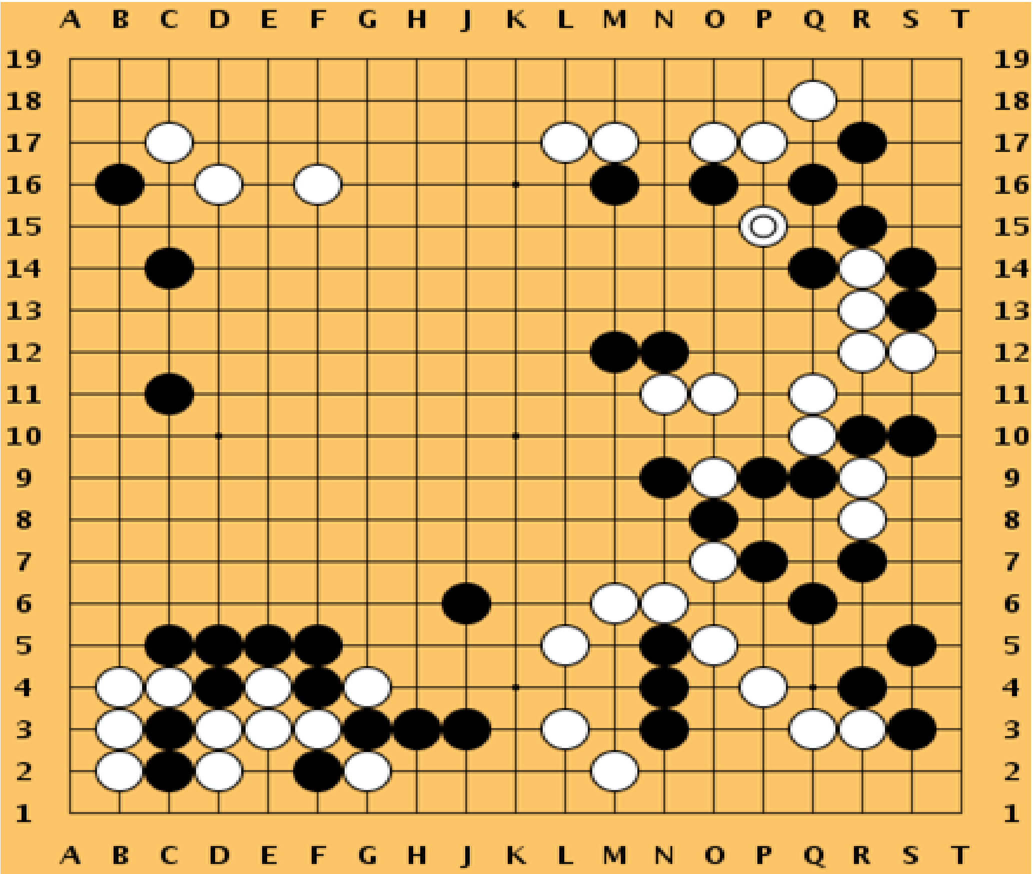
Case Study: Playing Games

AlphaGo: (January 2016)
- Used imitation learning + tree search + RL
- Beat 18-time world champion Lee Sedol

AlphaGo Zero (October 2017)
- Simplified version of AlphaGo
- No longer using imitation learning
- Beat (at the time) #1 ranked Ke Jie

Alpha Zero (December 2018)
- Generalized to other games: Chess and Shogi

MuZero (November 2019)
- Plans through a learned model of the game



Silver et al, "Mastering the game of Go with deep neural networks and tree search", Nature 2016
Silver et al, "Mastering the game of Go without human knowledge", Nature 2017
Silver et al, "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play", Science 2018
Schrittwieser et al, "Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model", arXiv 2019

Case Study: Playing Games

AlphaGo: (January 2016)
- Used imitation learning + tree search + RL
- Beat 18-time world champion Lee Sedol

AlphaGo Zero (October 2017)
- Simplified version of AlphaGo
- No longer using imitation learning
- Beat (at the time) #1 ranked Ke Jie

Alpha Zero (December 2018)
- Generalized to other games: Chess and Shogi

MuZero (November 2019)
- Plans through a learned model of the game

November 2019: Lee Sedol announces retirement

"With the debut of AI in Go games, I've realized that I'm not at the top even if I become the number one through frantic efforts"
"Even if I become the number one, there is an entity that cannot be defeated"

Silver et al, "Mastering the game of Go with deep neural networks and tree search", Nature 2016
Silver et al, "Mastering the game of Go without human knowledge", Nature 2017
Silver et al, "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play", Science 2018
Schrittwieser et al, "Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model", arXiv 2019

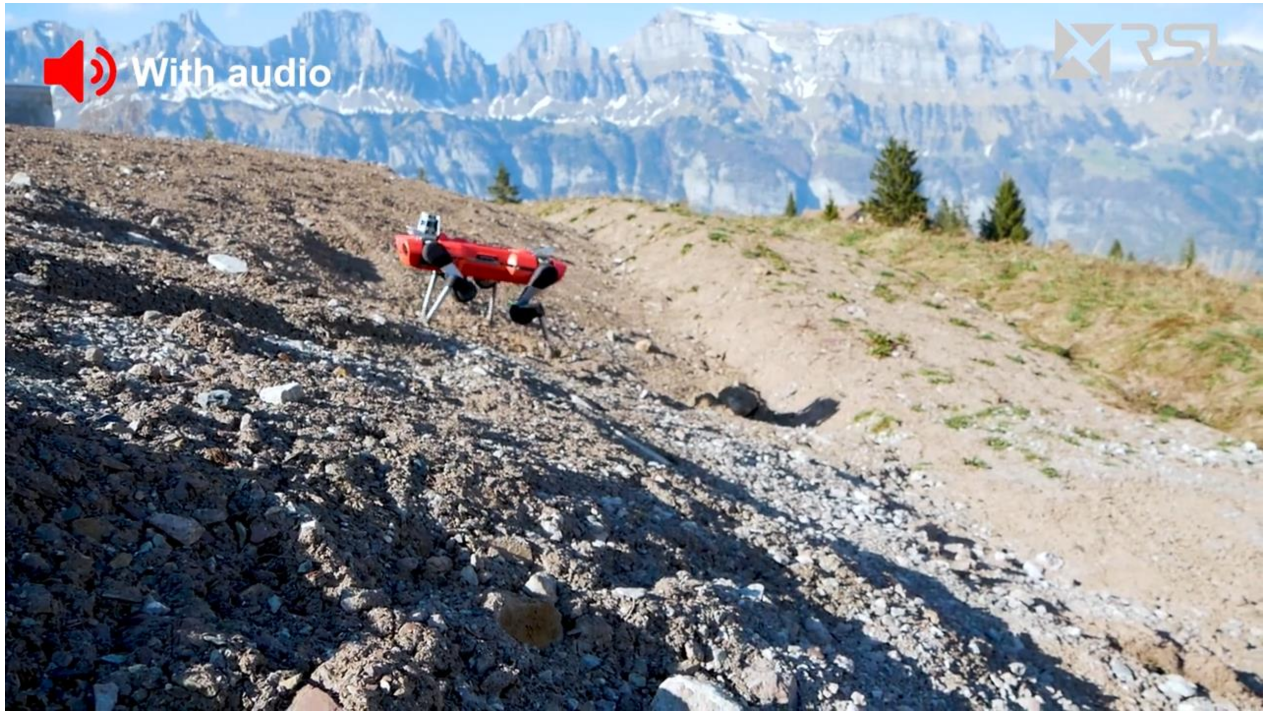
Quotes from: https://en.yna.co.kr/view/AEN20191127004800315

More Complex Games

StarCraft II: AlphaStar (October 2019)
Vinyals et al, “Grandmaster level in StarCraft II using multi-agent reinforcement learning”, Science 2018

Dota 2: OpenAI Five (April 2019)
No paper, only a blog post:
https://openai.com/five/#how-openai-five-works

In Robotics: Locomotion

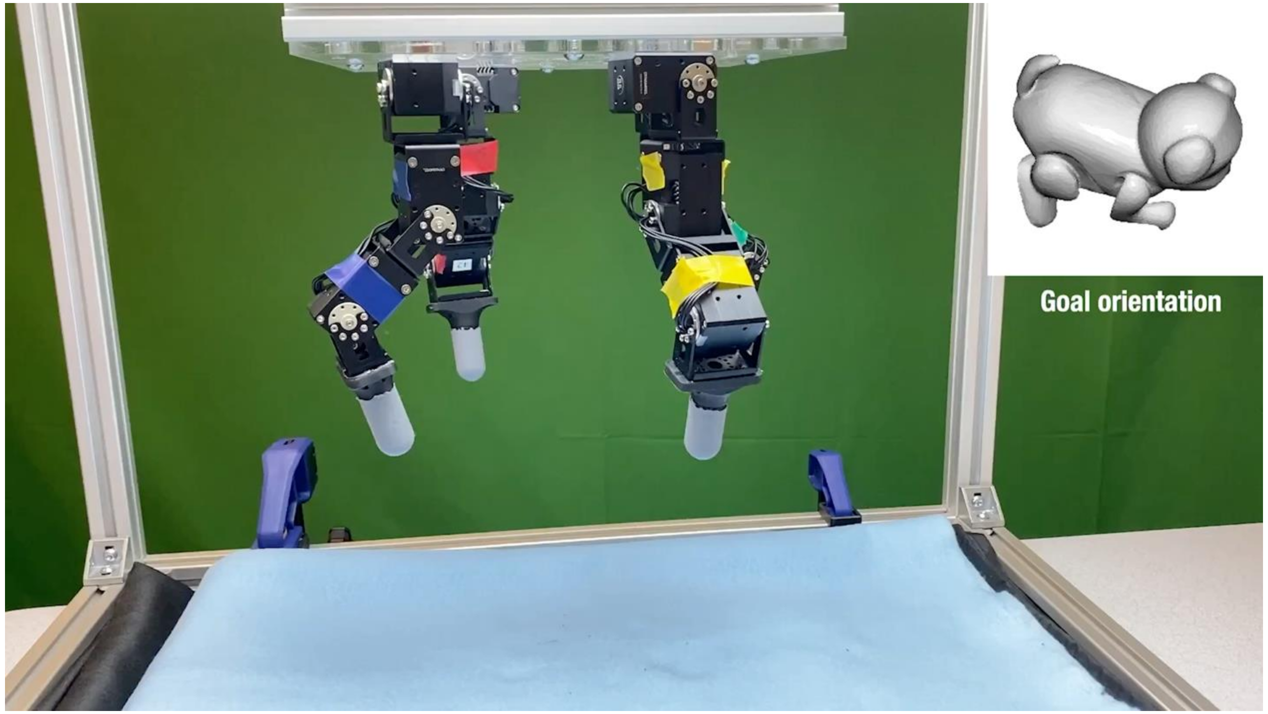


Learning Quadrupedal Locomotion over Challenging Terrain
Science Robotics 2020

Unitree, Dec. 2024

In Robotics: Dexterous Manipulation

Solving Rubik's Cube with a Robot Hand
OpenAI 2019



Visual Dexterity: In-Hand Reorientation of Novel and Complex Object Shapes, Science Robotics 2023

Problems of Model-Free RL

- Learns from trial and error
- Require extensive interactions

AlphaGo Zero: Google DeepMind supercomputer learns 3,000 years of human knowledge in 40 days

Problems of Model-Free RL

- Learns from trial and error
- Require extensive interactions

- Safety concerns
- Limited interpretability
  - What if things go wrong?

Problems of (Model-Free) RL

- Learns from trial and error
- Require extensive interactions

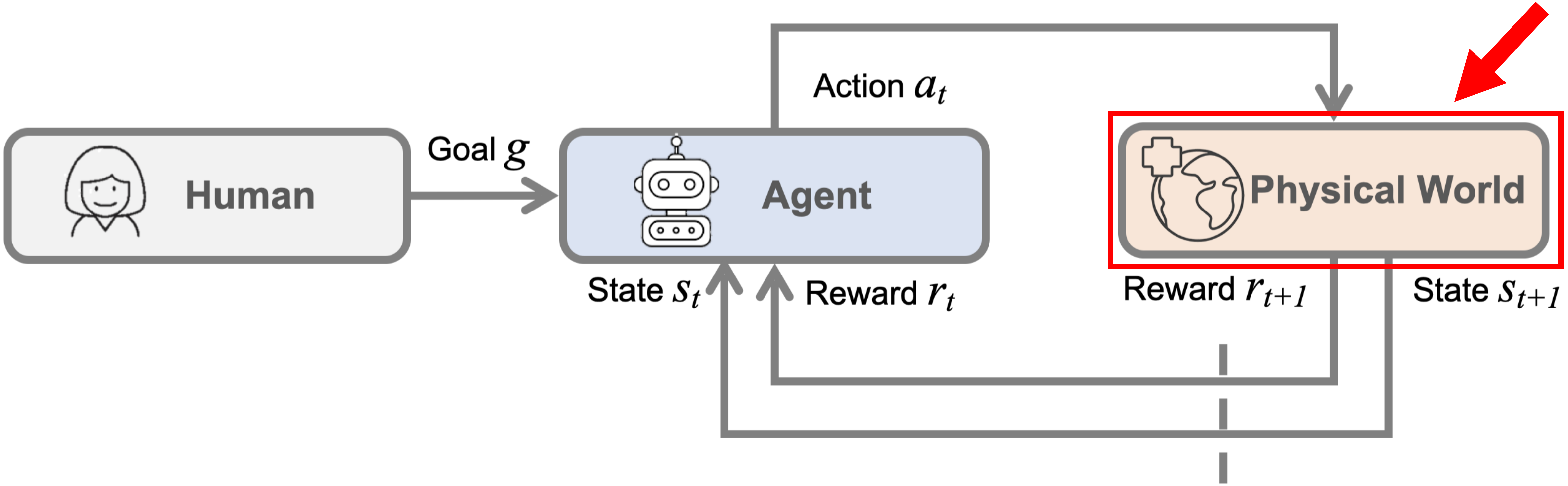
- Safety concerns
- Limited interpretability
  - What if things go wrong?

- Humans maintain an intuitive model of the world
  - Widely applicable
  - Sample efficient

Overview

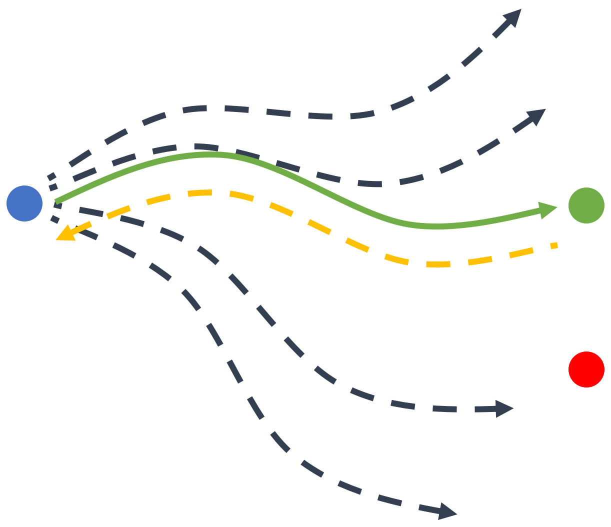
- Problem formulation
- Robot perception
- Reinforcement learning
- Model learning & model-based planning
- Imitation learning
- Robotic foundation models
- Remaining challenges

Model Learning & Model-Based Planning



Model Learning & Model-Based Planning

Learn a model of the world's state transition function $P(s_{t+1}|s_t, a_t)$ and then use planning through the model to make decisions



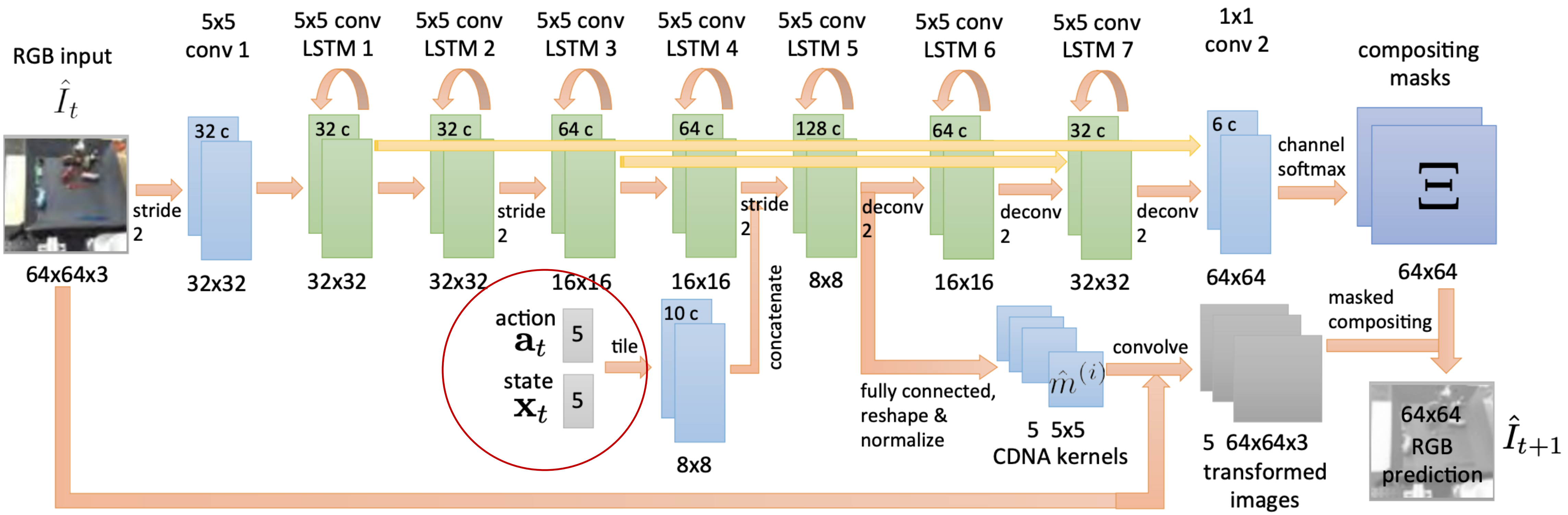
Model might not be accurate enough.

1. Execute the first action
2. Obtain new state
3. Re-optimize the action sequence using gradient descent

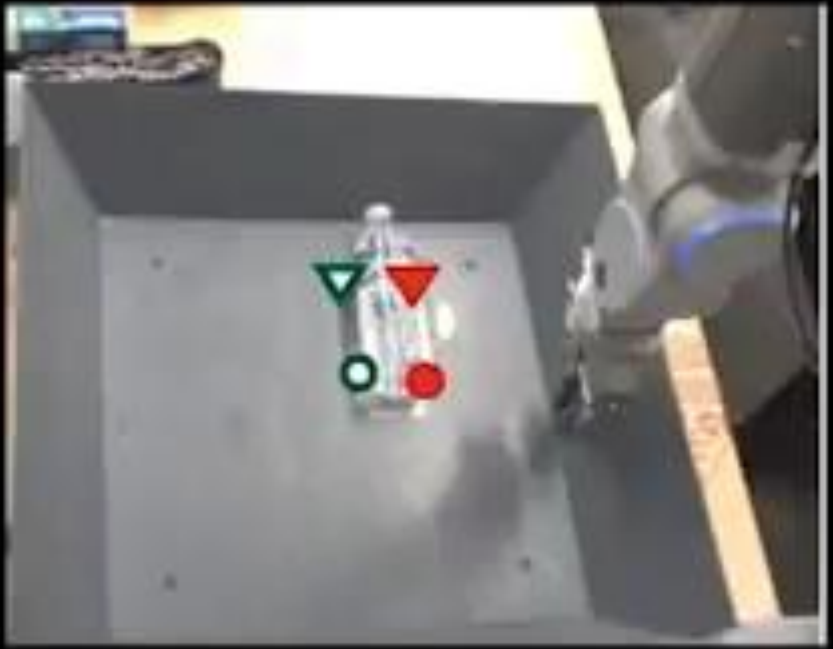
Key: GPU for parallel sampling / gradient descent

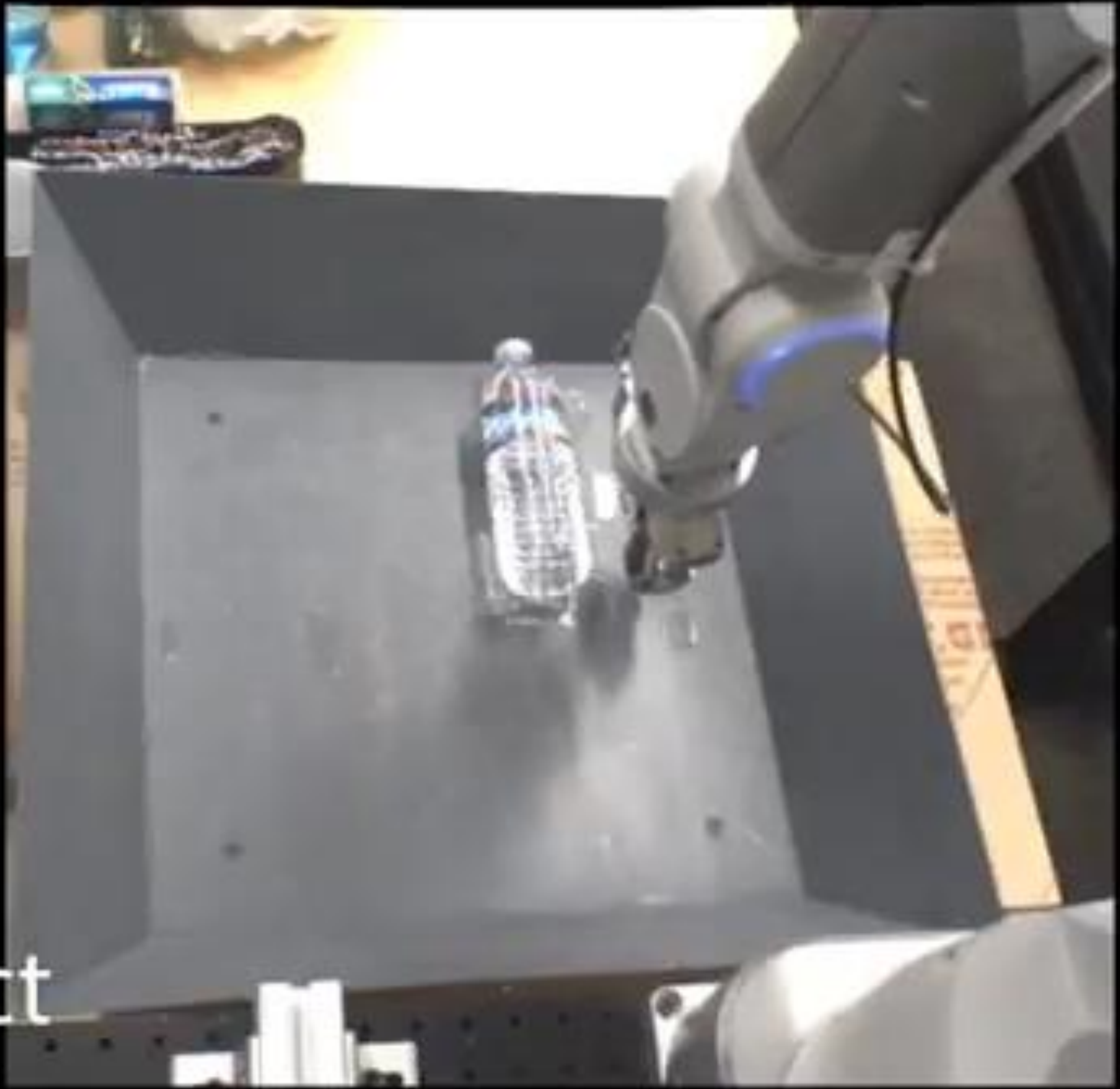
Key question: what should be the form of s_t?

Pixel Dynamics - Deep Visual Foresight



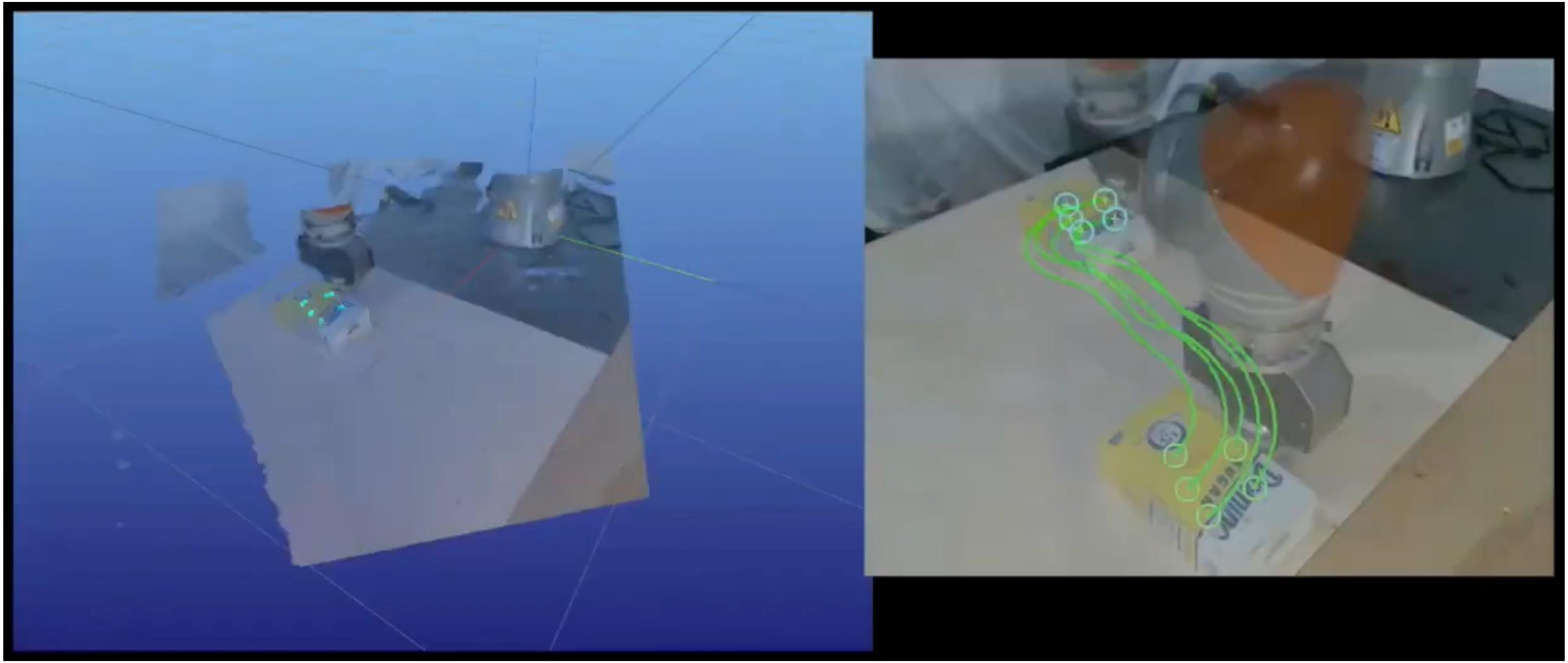
Finn and Levine, "Deep Visual Foresight for Planning Robot Motion", ICRA 2017



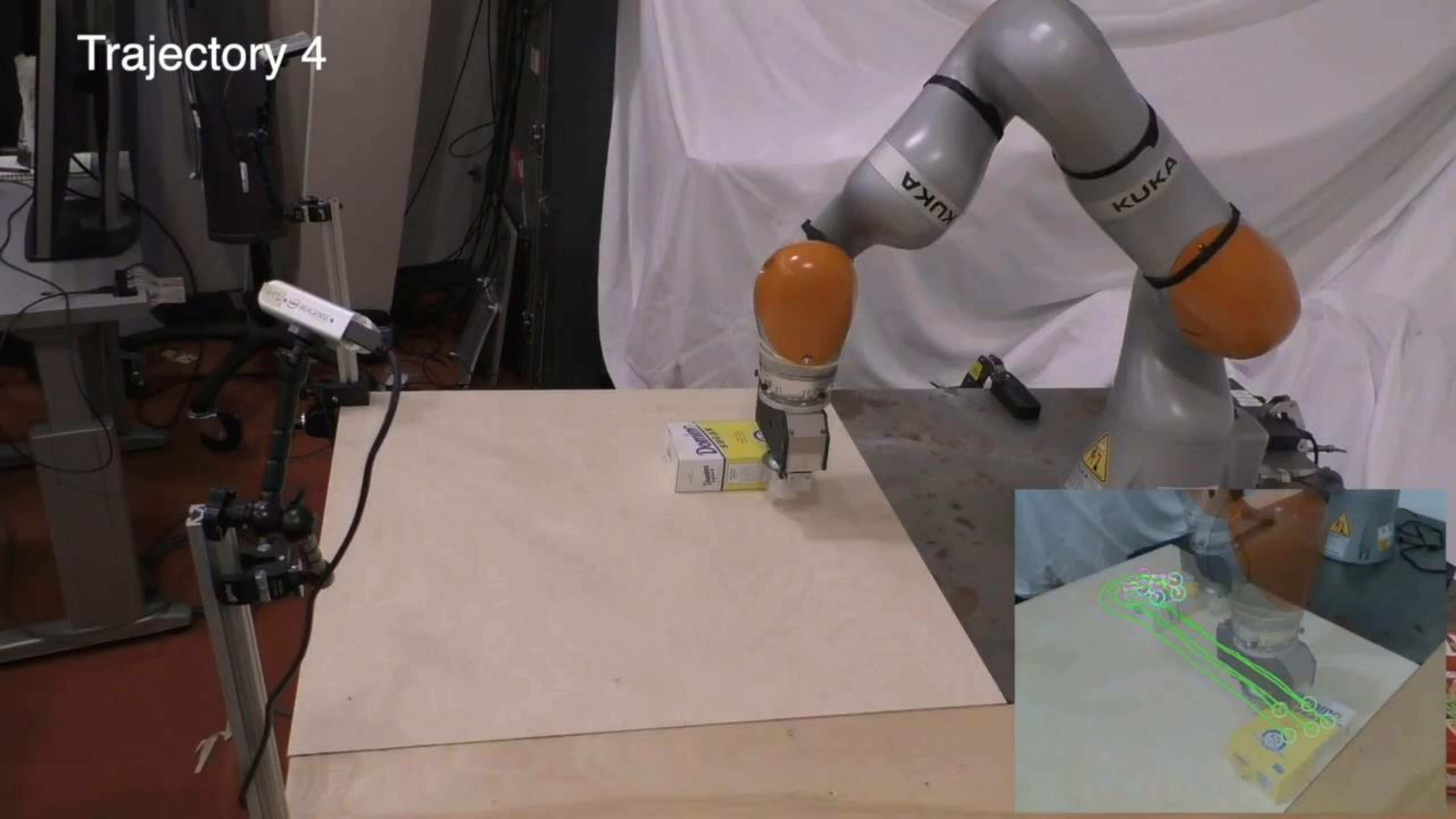


transparent object

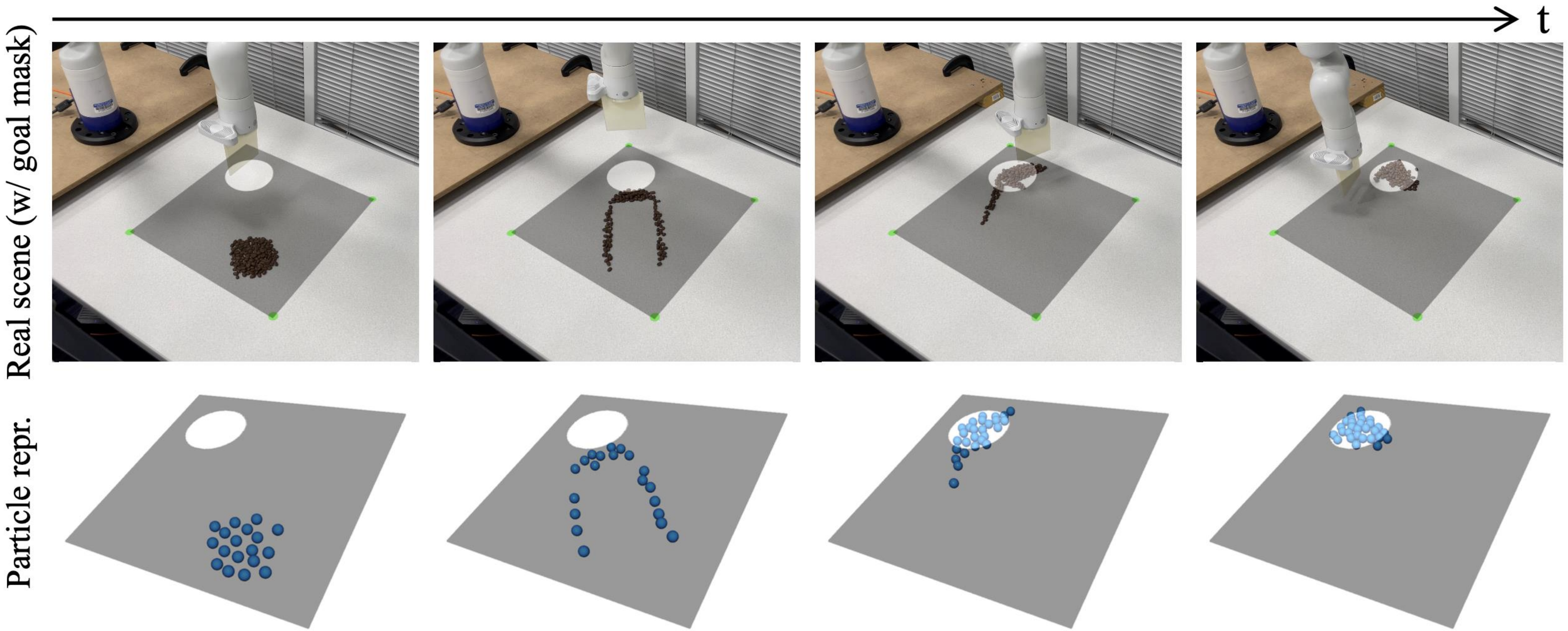
Keypoint Dynamics



Manuelli, Li, Florence, Tedrake, “Keypoints into the Future: Self-Supervised Correspondence in Model-Based Reinforcement Learning”, CoRL 2020

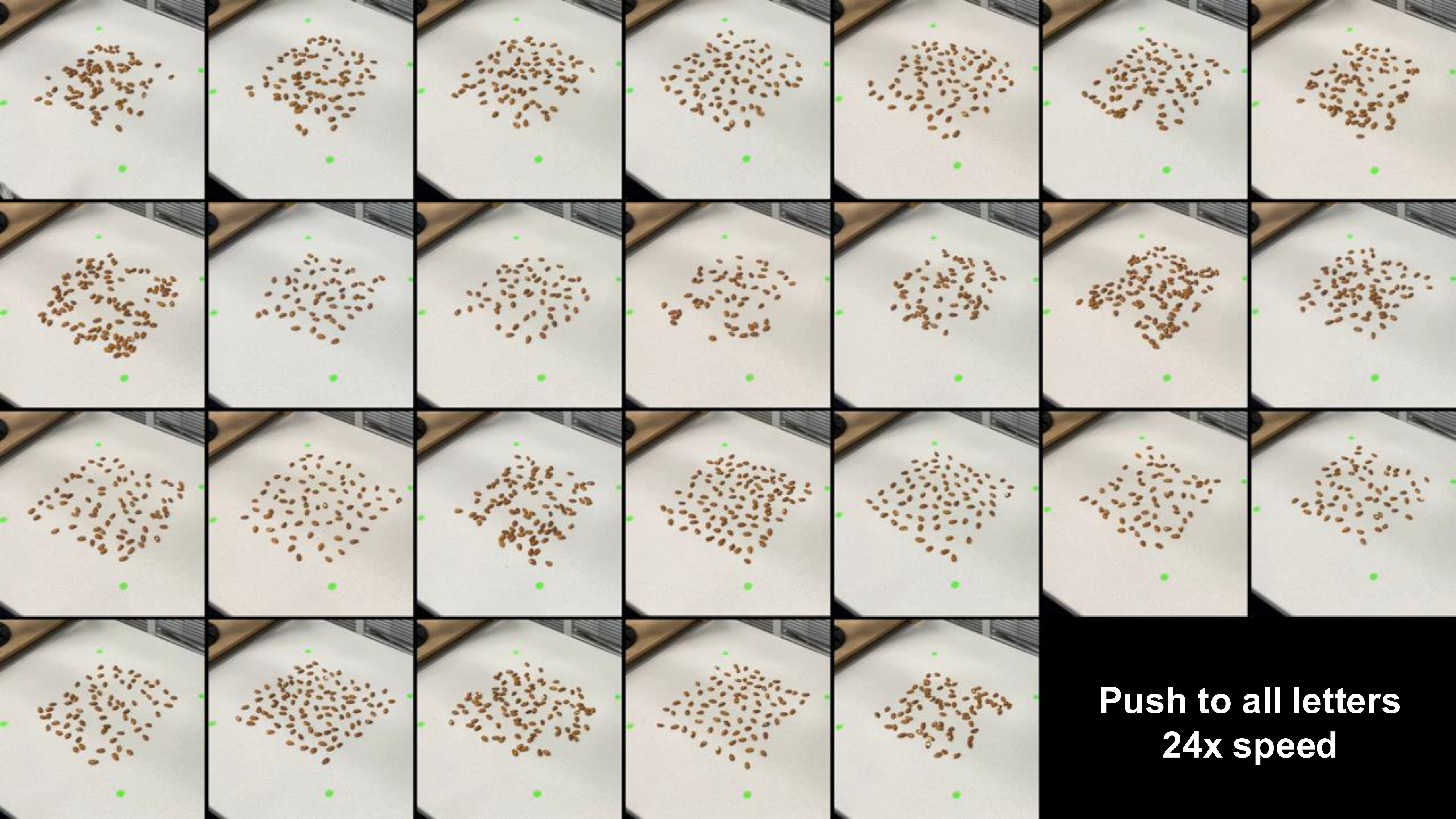

Trajectory 4
KUKA
KUKA

Particle Dynamics



Wang, Li, Driggs-Campbell, Fei-Fei, Wu, "Dynamic-Resolution Model Learning for Object Pile Manipulation", RSS 2023

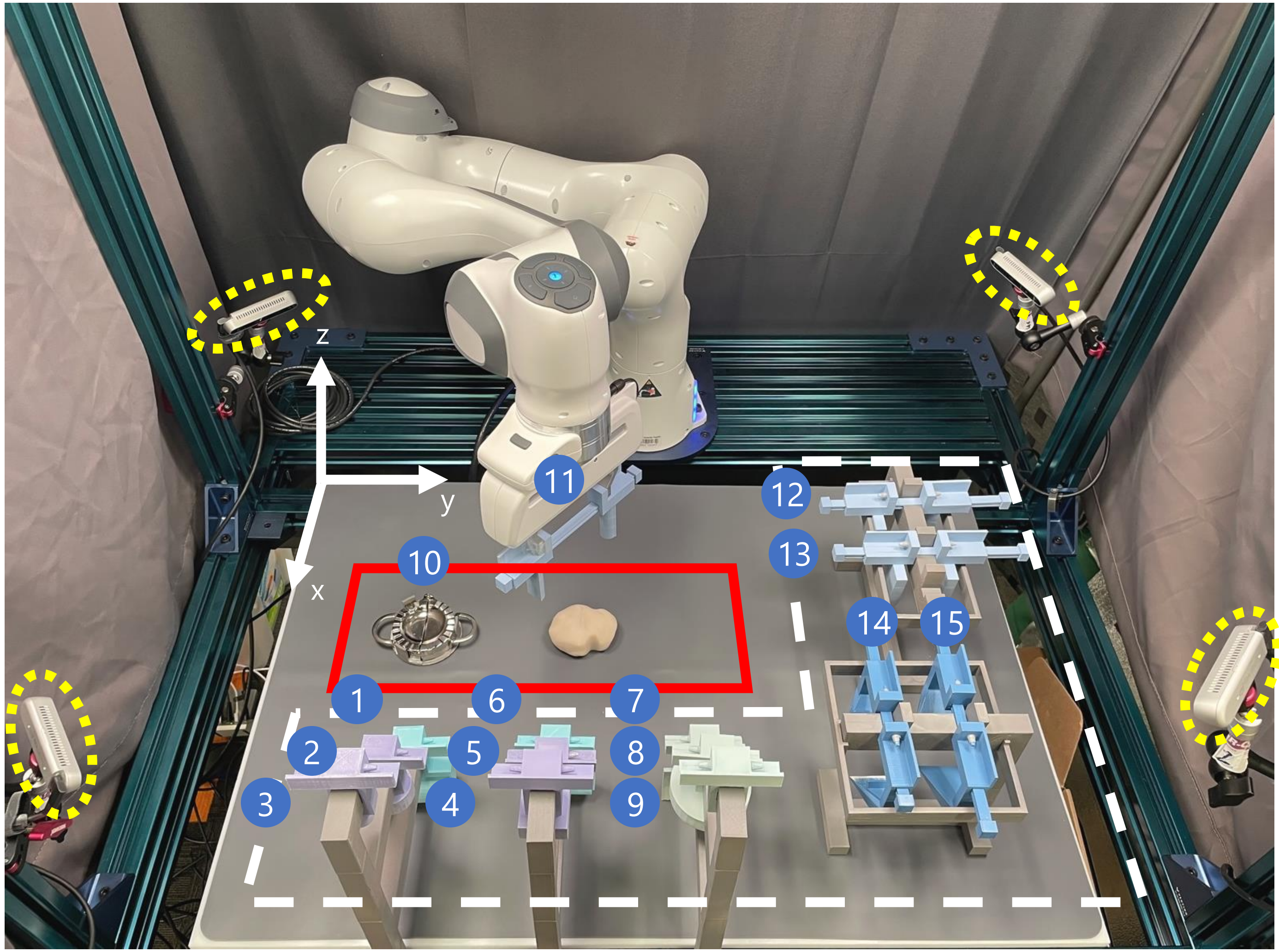

Granola
Rice
Carrot
Candy
24x speed



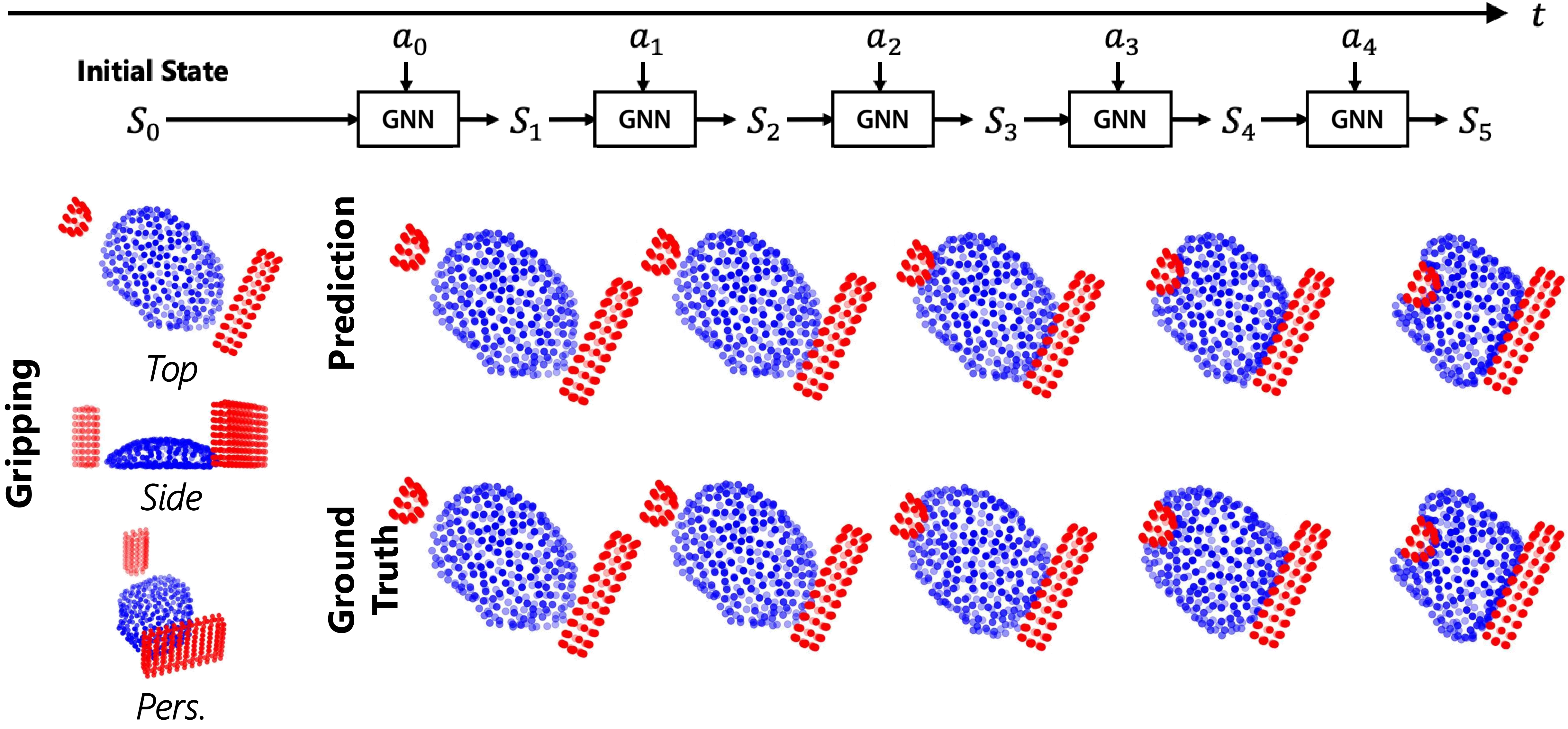
Push to all letters
24x speed

Particle Dynamics

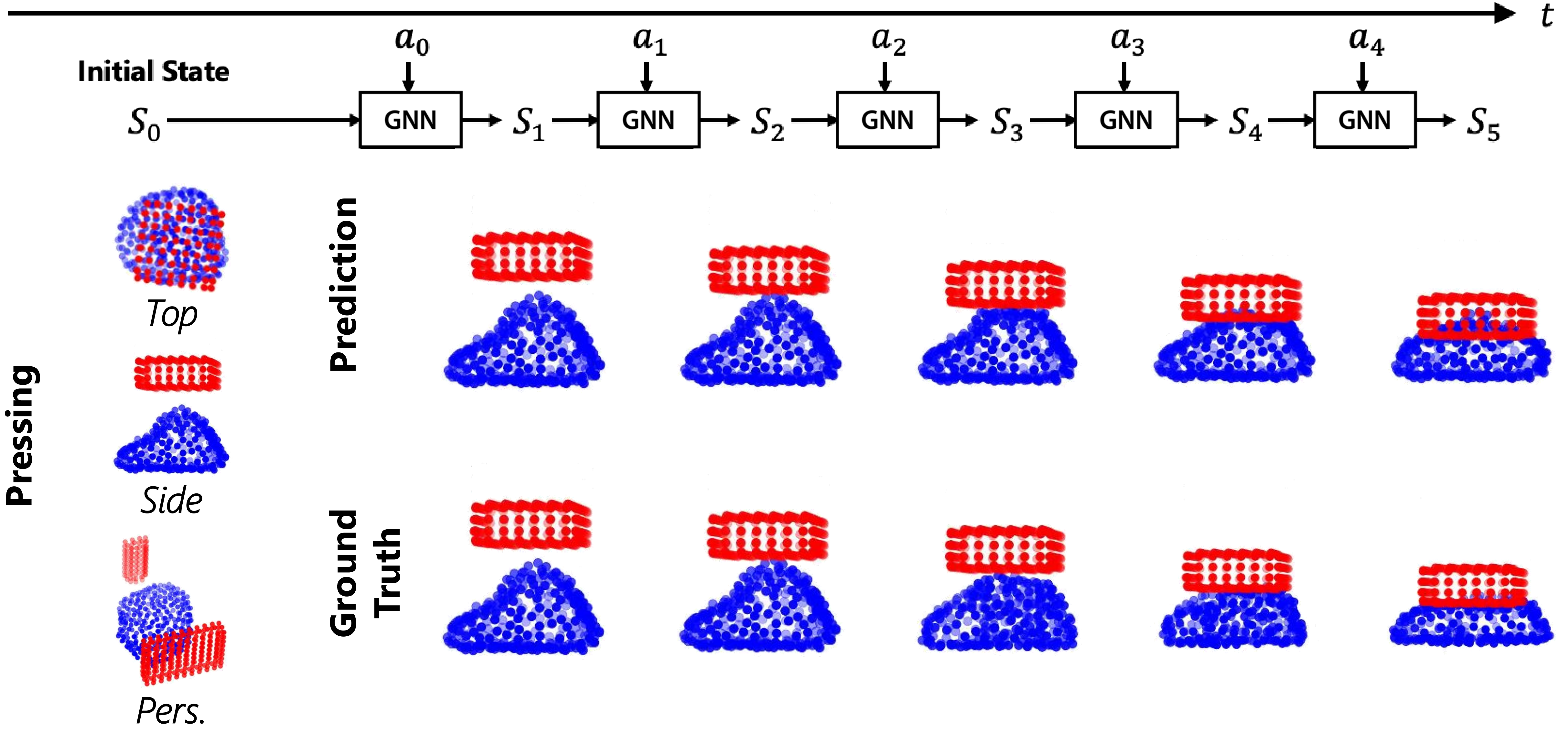
Haochen Shi*, Huazhe Xu*, Samuel Clarke, **Yunzhu Li**, and Jiajun Wu
RoboCook: Long-Horizon Elasto-Plastic Object Manipulation with Diverse Tools
Conference on Robot Learning (CoRL) 2023 – **Best Systems Paper Award**



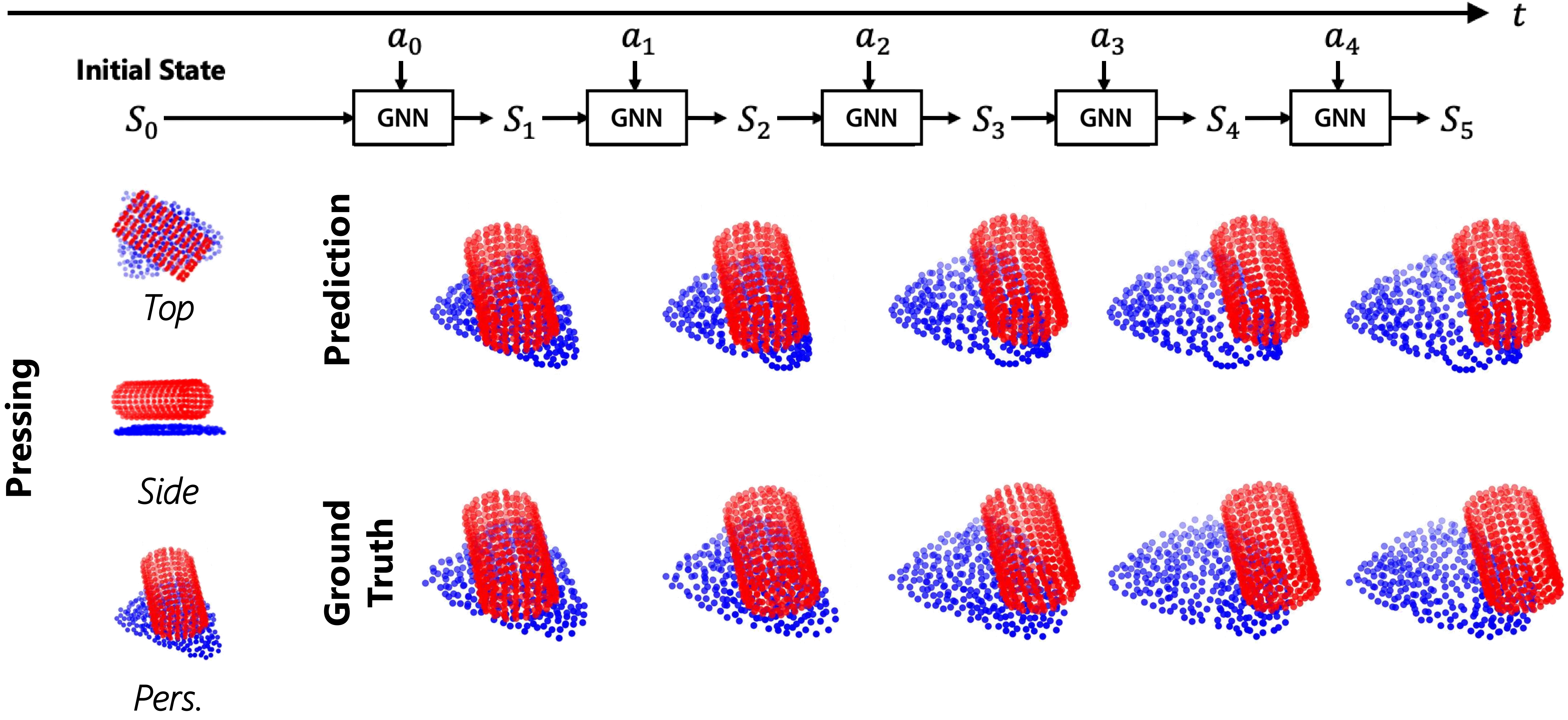
Particle Dynamics – Future Prediction

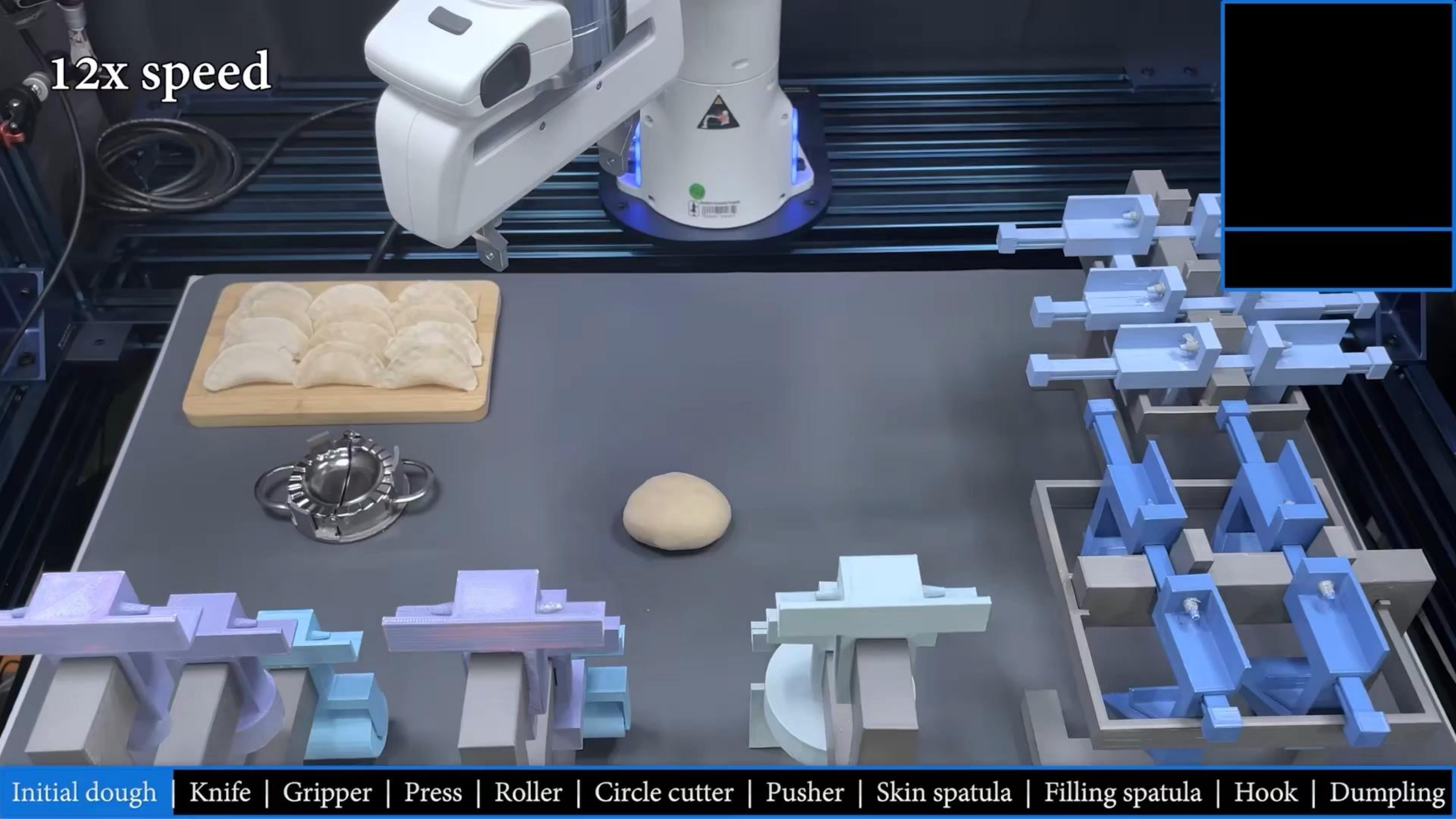


Particle Dynamics – Future Prediction



Particle Dynamics – Future Prediction



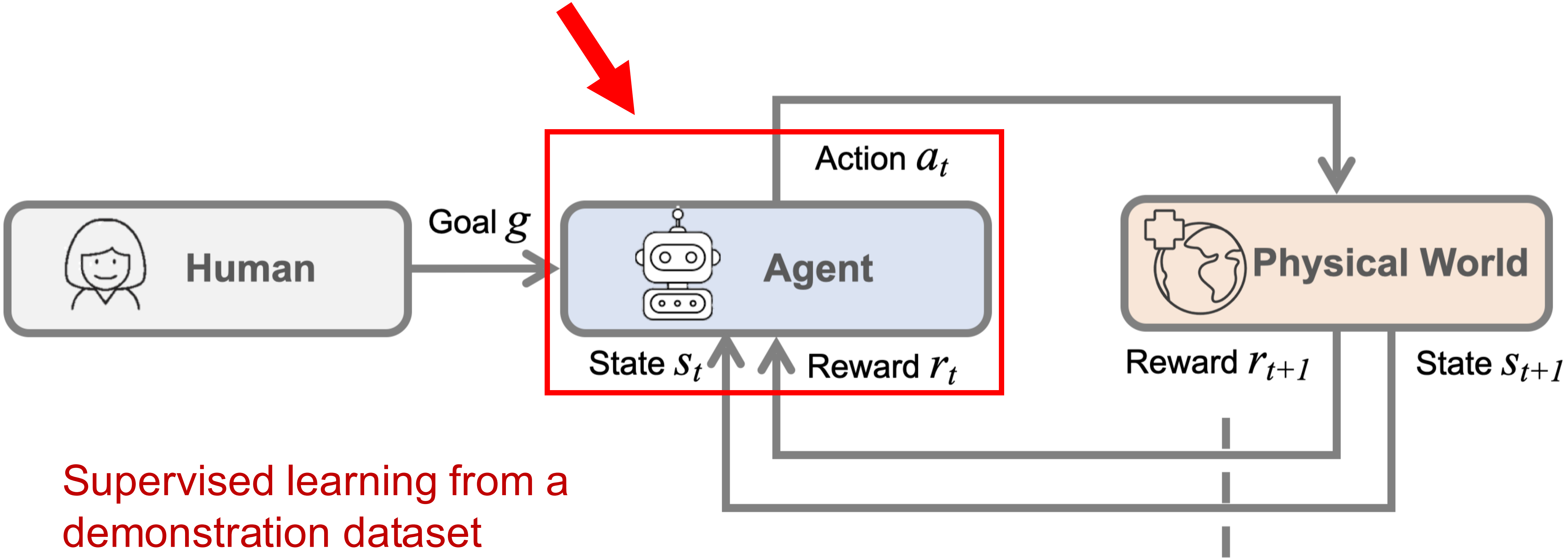


12x speed
Initial dough | Knife | Gripper | Press | Roller | Circle cutter | Pusher | Skin spatula | Filling spatula | Hook | Dumpling

Overview

- Problem formulation
- Robot perception
- Reinforcement learning
- Model learning & model-based planning
- Imitation learning
- Robotic foundation models
- Remaining challenges

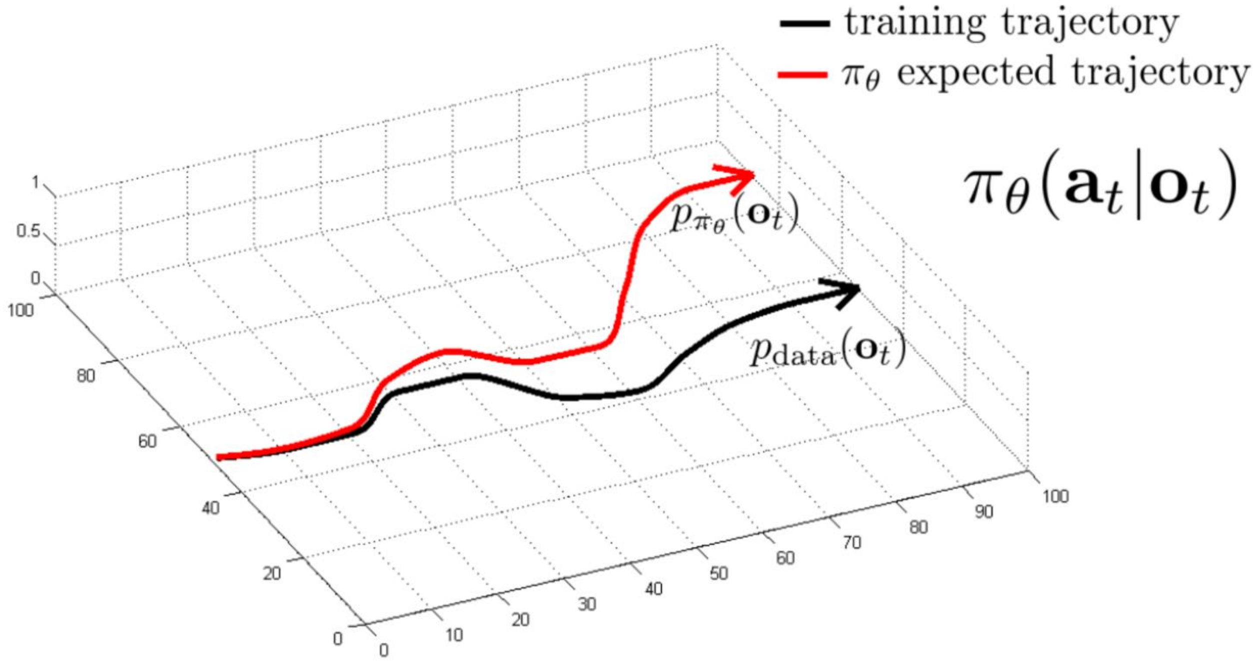
Imitation Learning



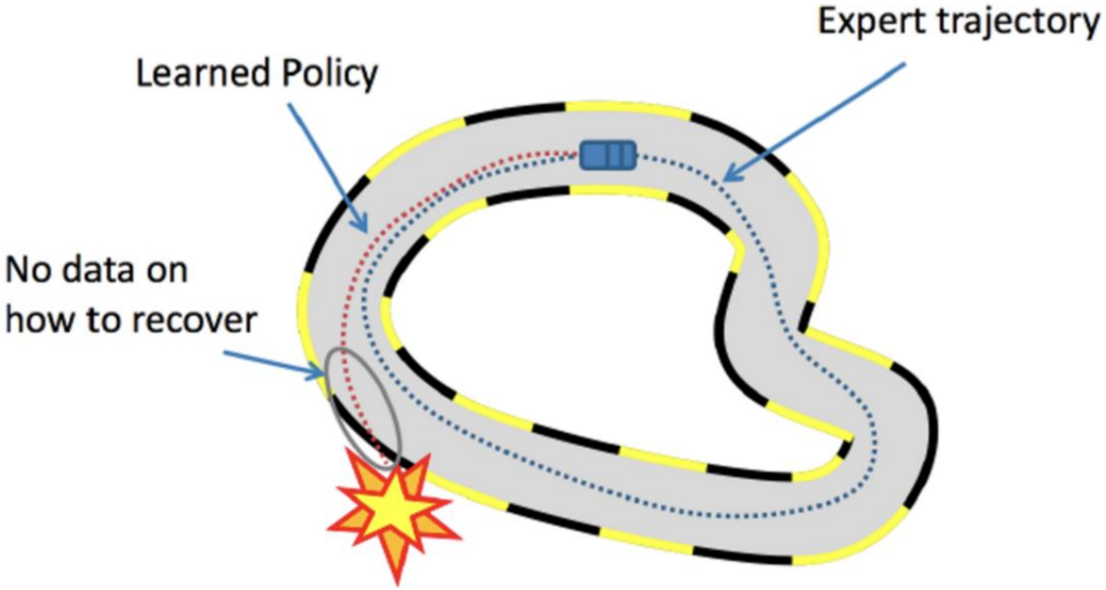
Supervised learning from a demonstration dataset

Learning from Demonstrations

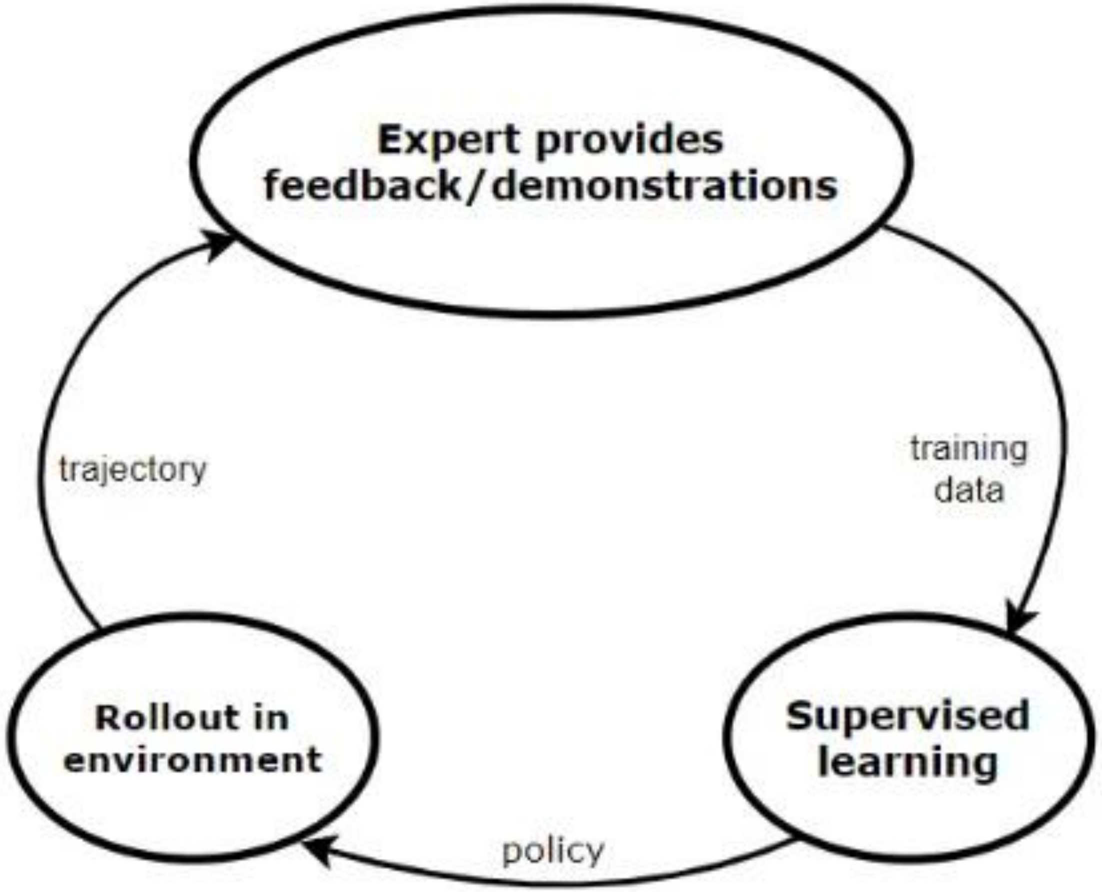
Behavior Cloning (BC)



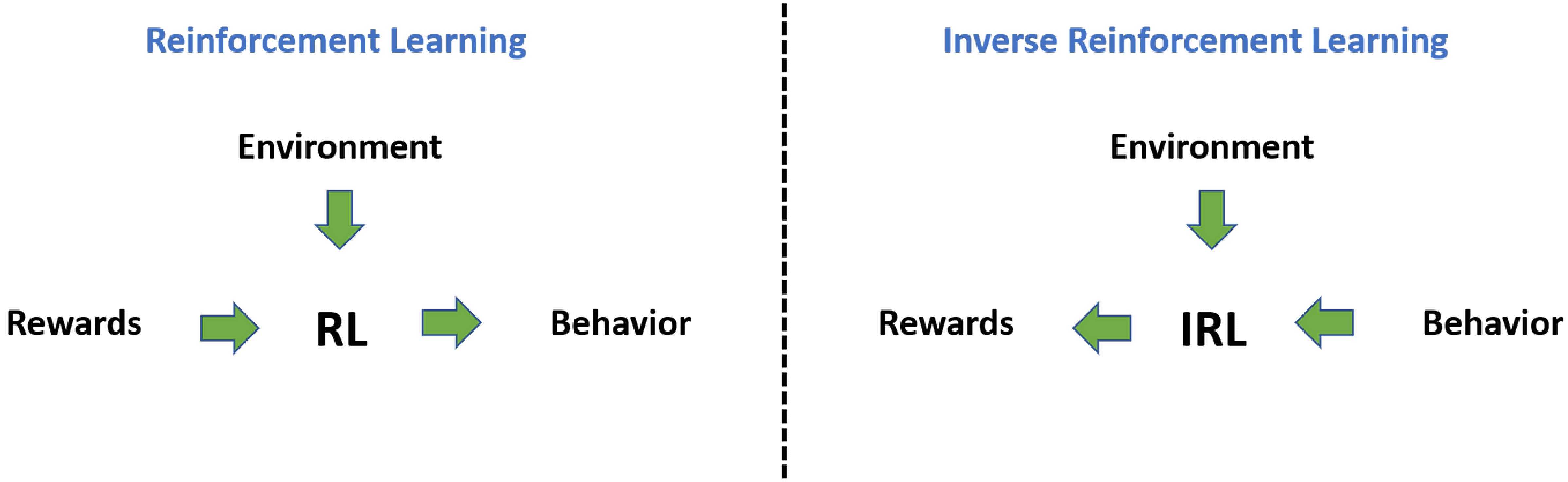
$$\pi_\theta(\mathbf{a}_t|\mathbf{o}_t)$$



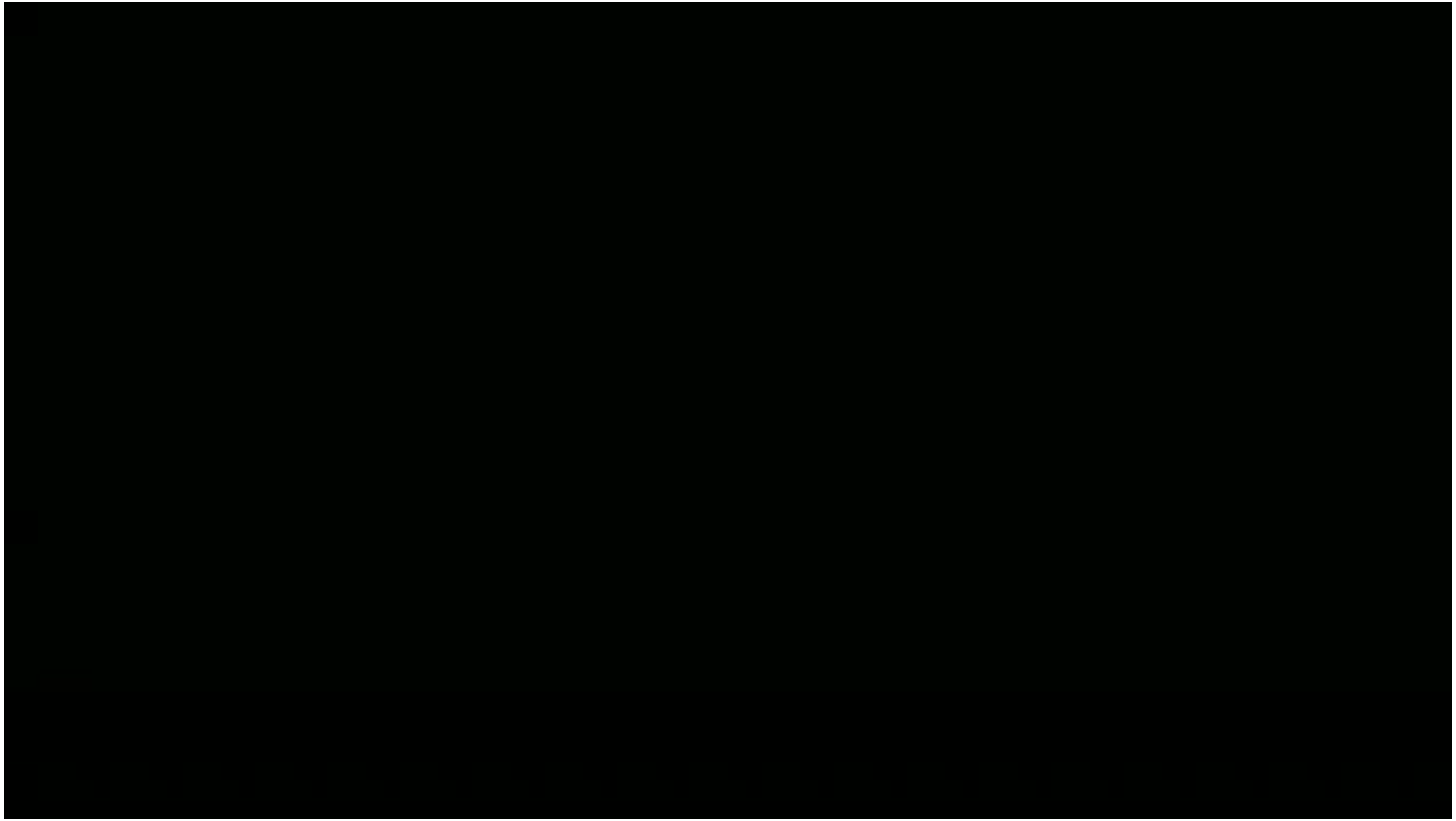
Iterative Collection of Expert Demonstrations



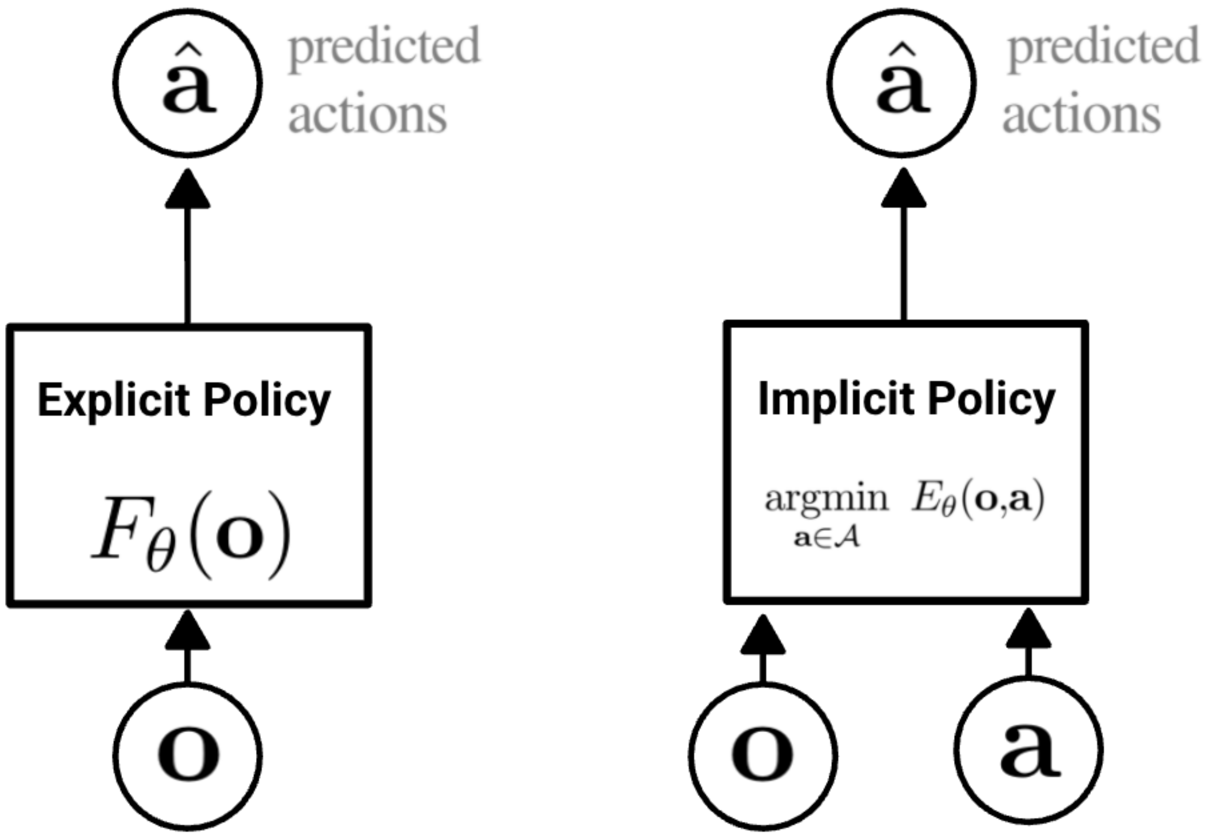
Inverse Reinforcement Learning (IRL)



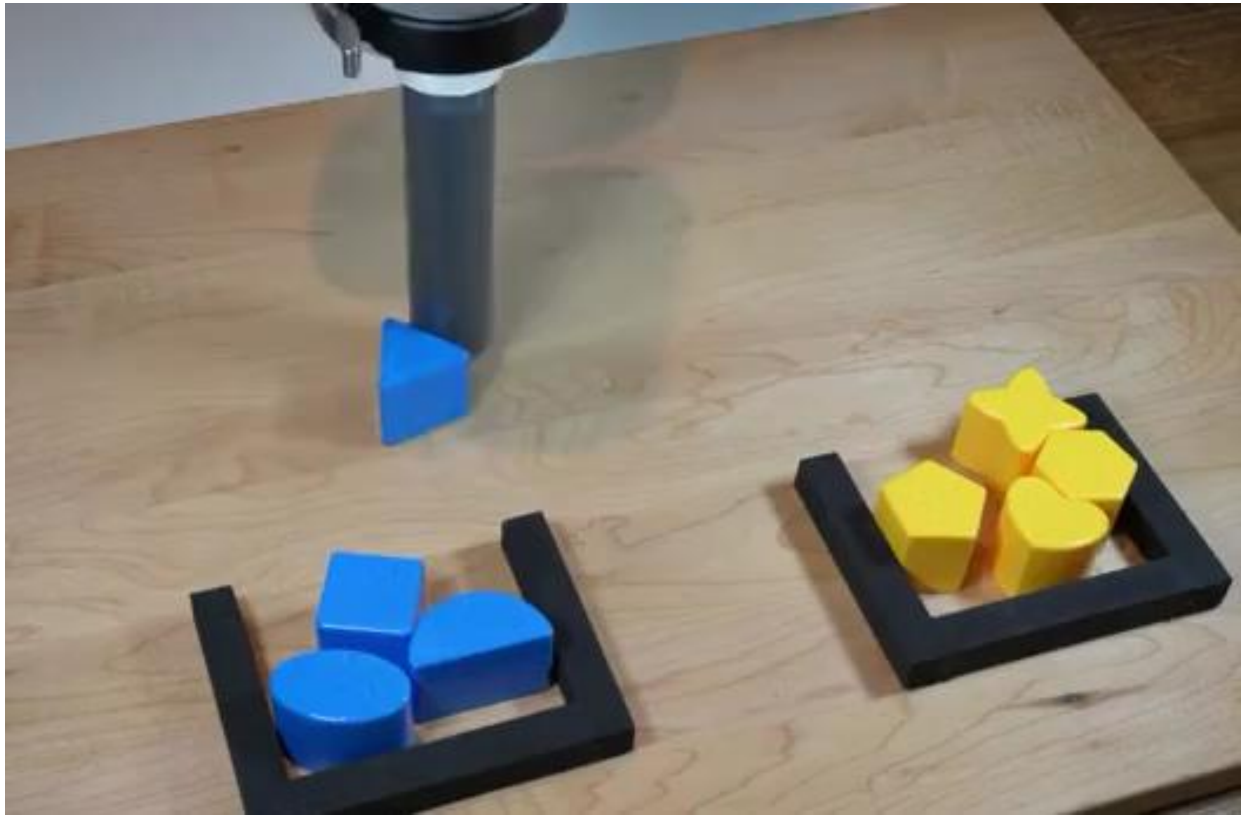
Inverse Reinforcement Learning (IRL)

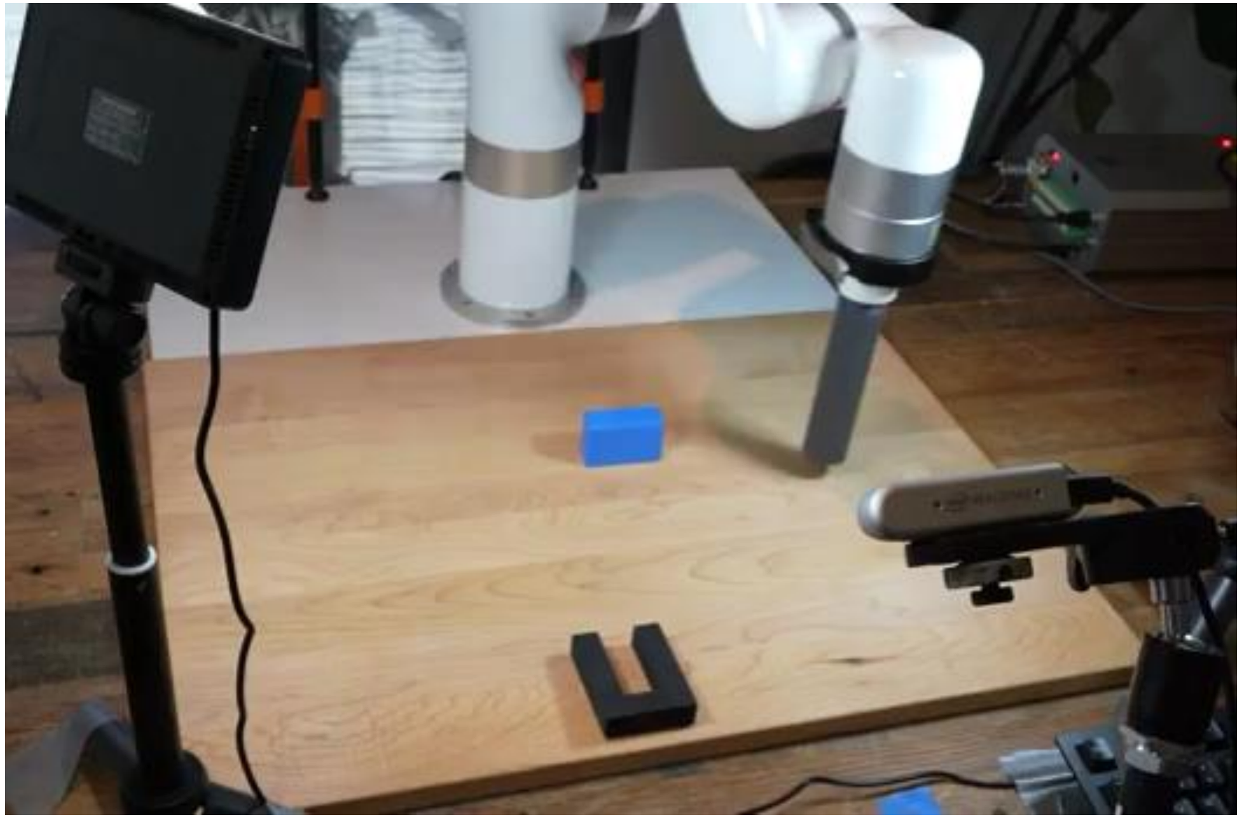


Implicit Behavior Cloning (IBC)



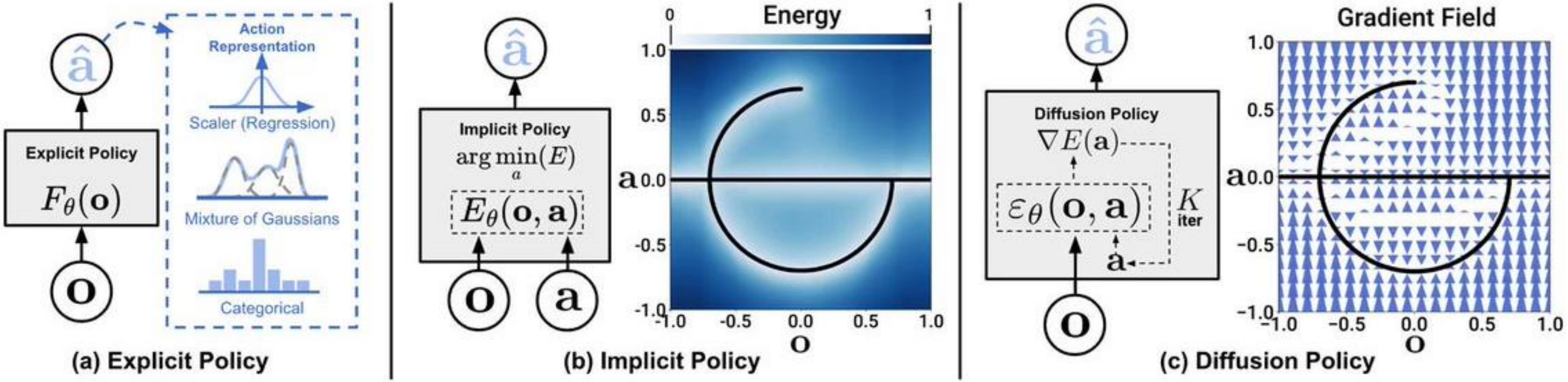
Implicit Behavior Cloning (IBC)





Diffusion Policies

Visuomotor Policy Learning via Action Diffusion



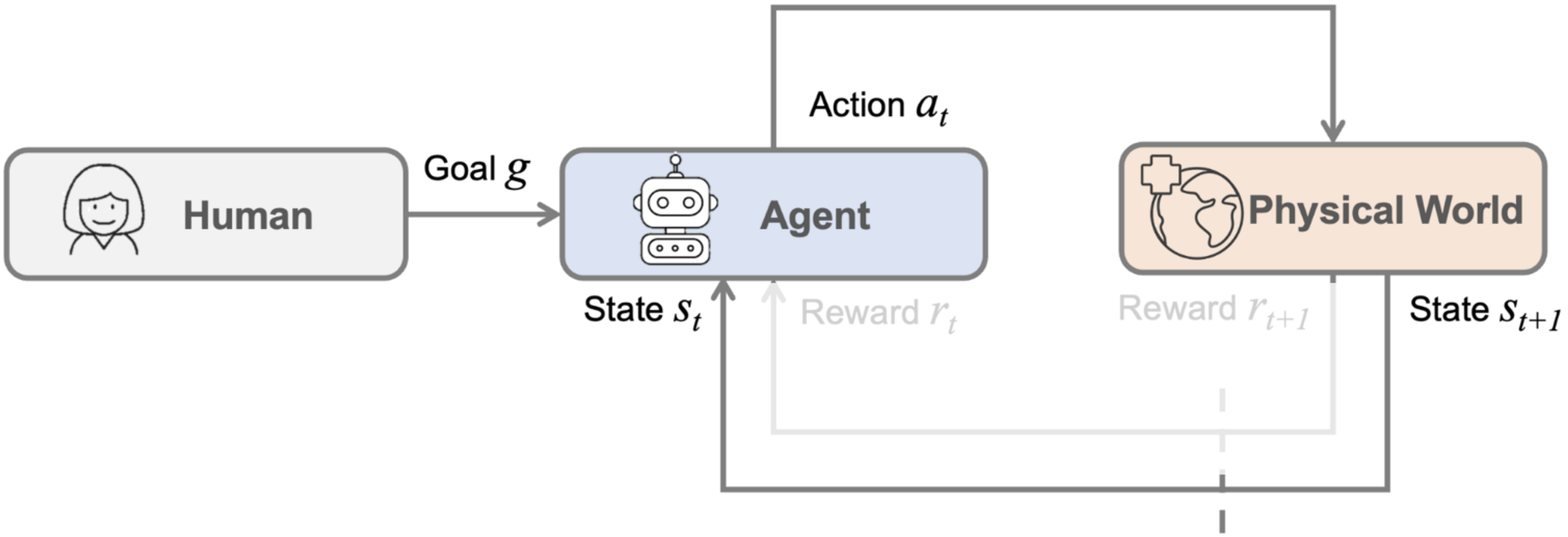
(a) Explicit Policy (b) Implicit Policy (c) Diffusion Policy

Overview

- Problem formulation
- Robot perception
- Reinforcement learning
- Model learning & model-based planning
- Imitation learning
- Robotic foundation models
- Remaining challenges

Robotic Foundation Models

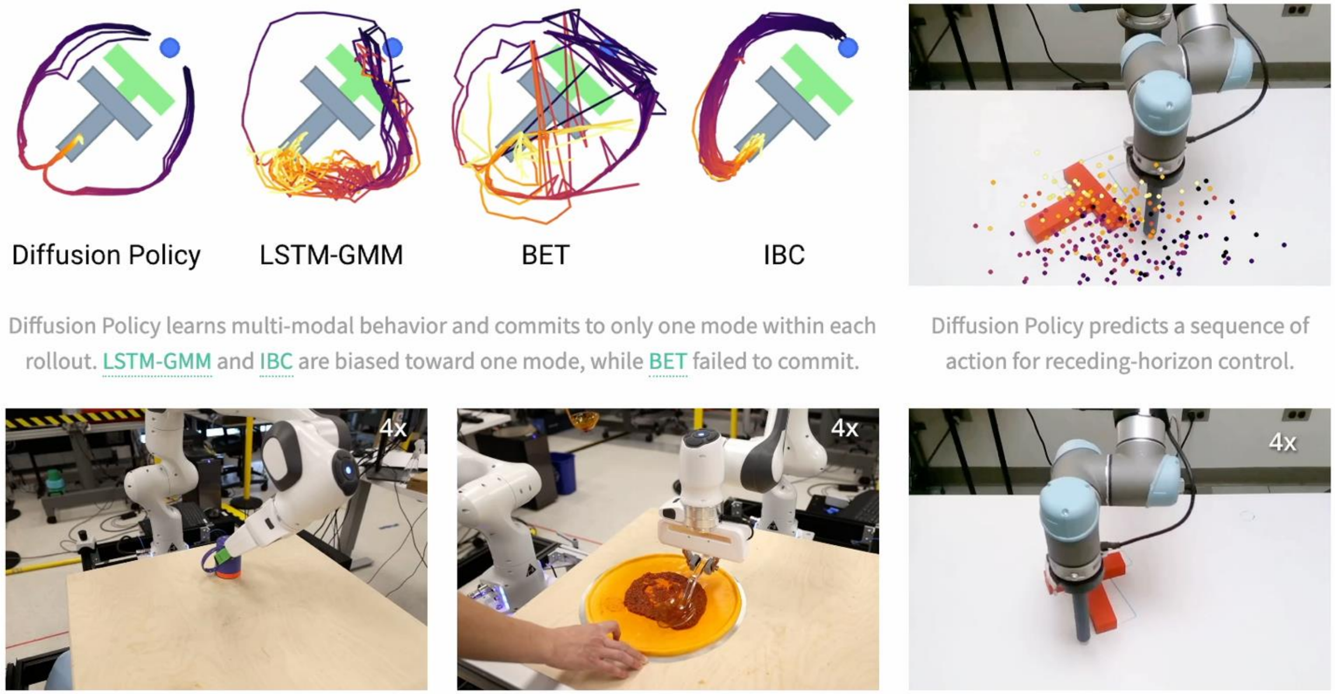
- What is a Robotic Foundation Model?
  - No explicit representation of states / transition functions
  - A policy that maps (observation/state, goal) to action



Robotic Foundation Models

- What is a Robotic Foundation Model?
  - No explicit representation of states / transition functions
  - A policy that maps (observation/state, goal) to action

Imitation Learning
(Chi et al., Diffusion Policy)



Reinforcement Learning
(OpenAI, Solving Rubik's Cube)

Robotic Foundation Models

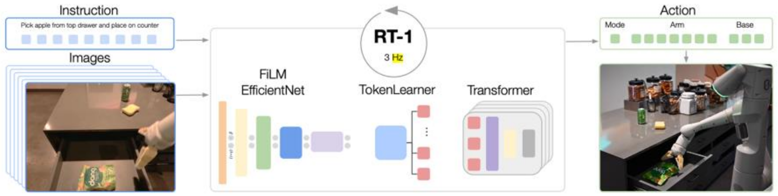
- What is a Robotic Foundation Model?
  - No explicit representation of states / transition functions
  - A policy that maps (observation/state, goal) to action

- Current Foundational Vision-and-Language Models
  - The output may **not** always be **perfect**.
  - It will always generate something **reasonable**.

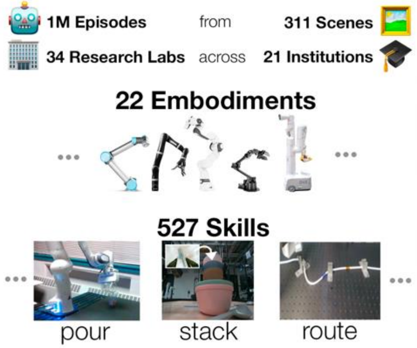
- Robotic Foundation Models
  - The synthesized action may **not** always be **optimal**.
  - The generated trajectory will always be **beautiful** and **reasonable**.

- Different names
  - Vision-Language-Action Models (VLAs), Large behavior models (LBMs)

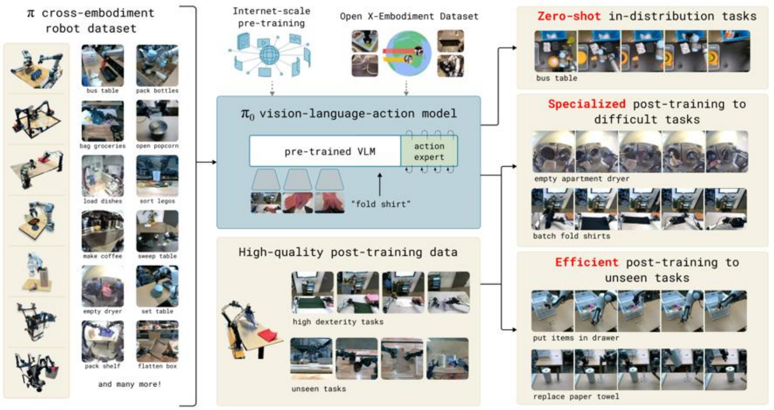
Robotic Foundation Models



RT-1 (Dec. 2022)

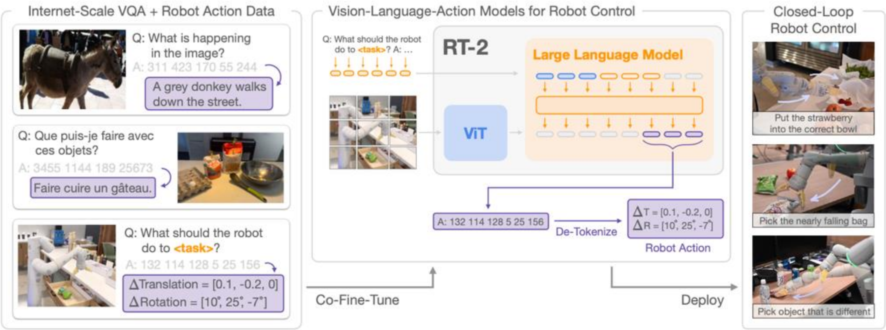


RT-X (Oct. 2023)

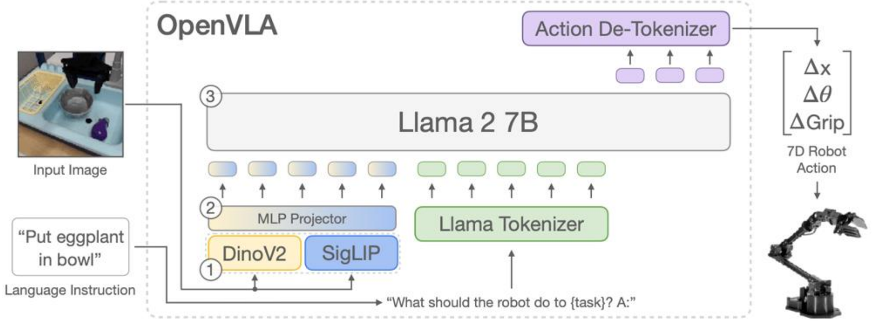


Pi-Zero (Oct. 2024)

RT-2 (Jul. 2023)



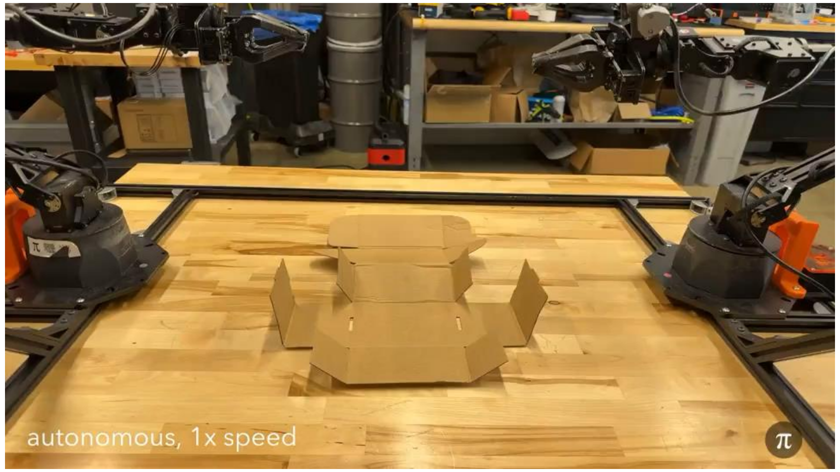
OpenVLA (Jun. 2024)

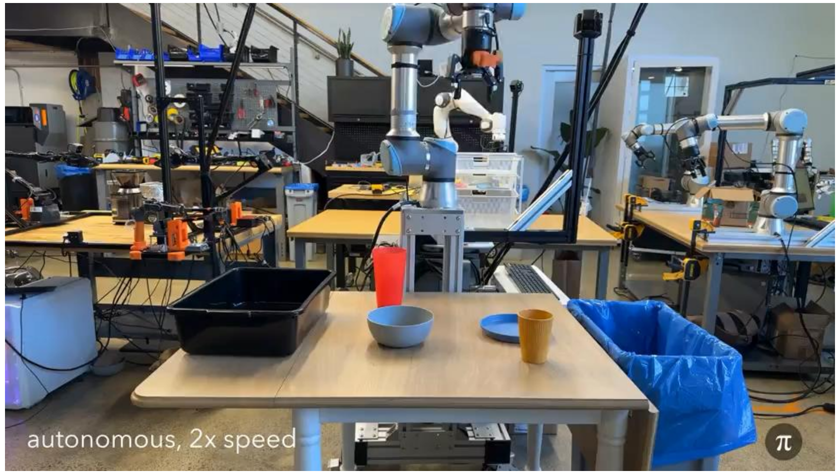


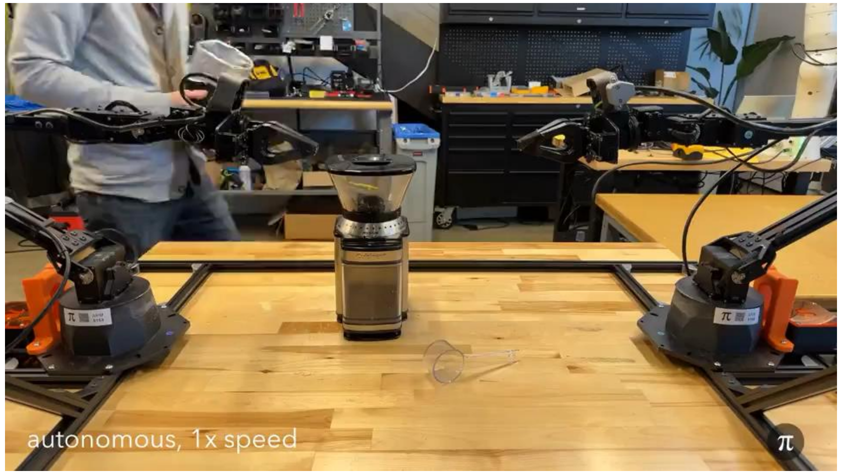
Helix (Figure)
Hi-Robot (PI)
Gemini Robotics
Pi-0.5 (PI)
GR00T (Nvidia)
DYNA-1
…

Pi-Zero by Physical Intelligence

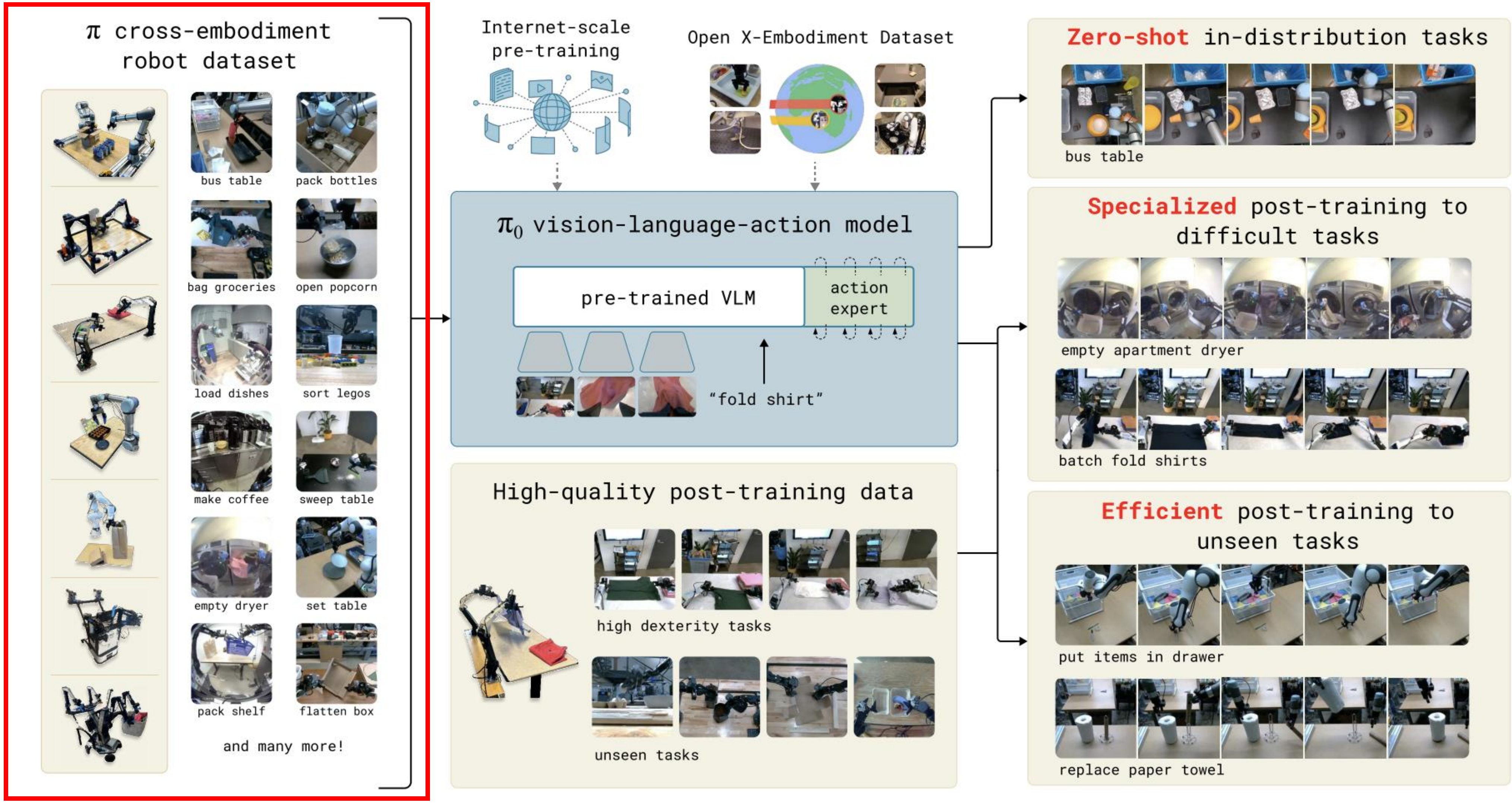
- First released in October 2024





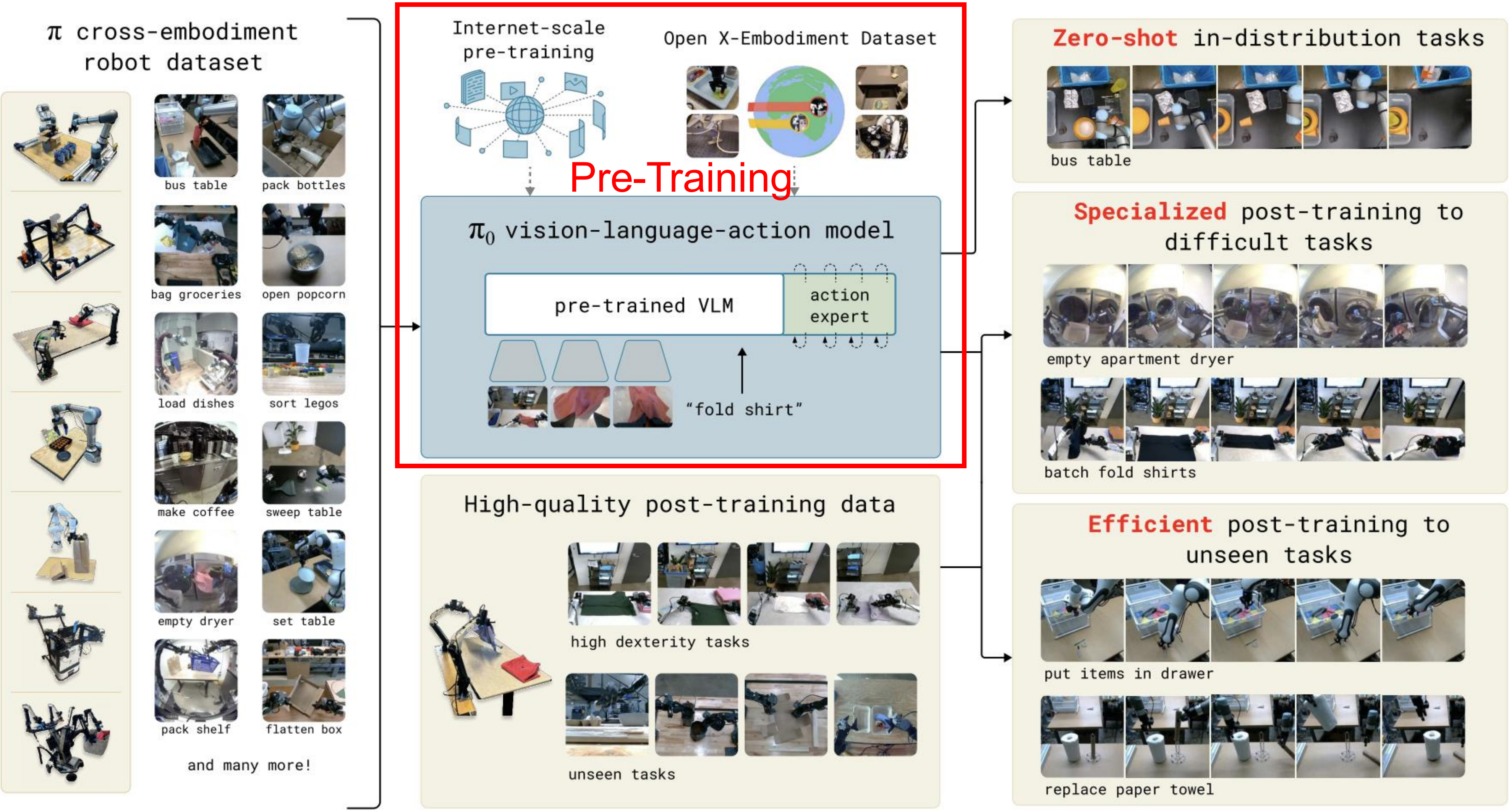


Pi-Zero by Physical Intelligence

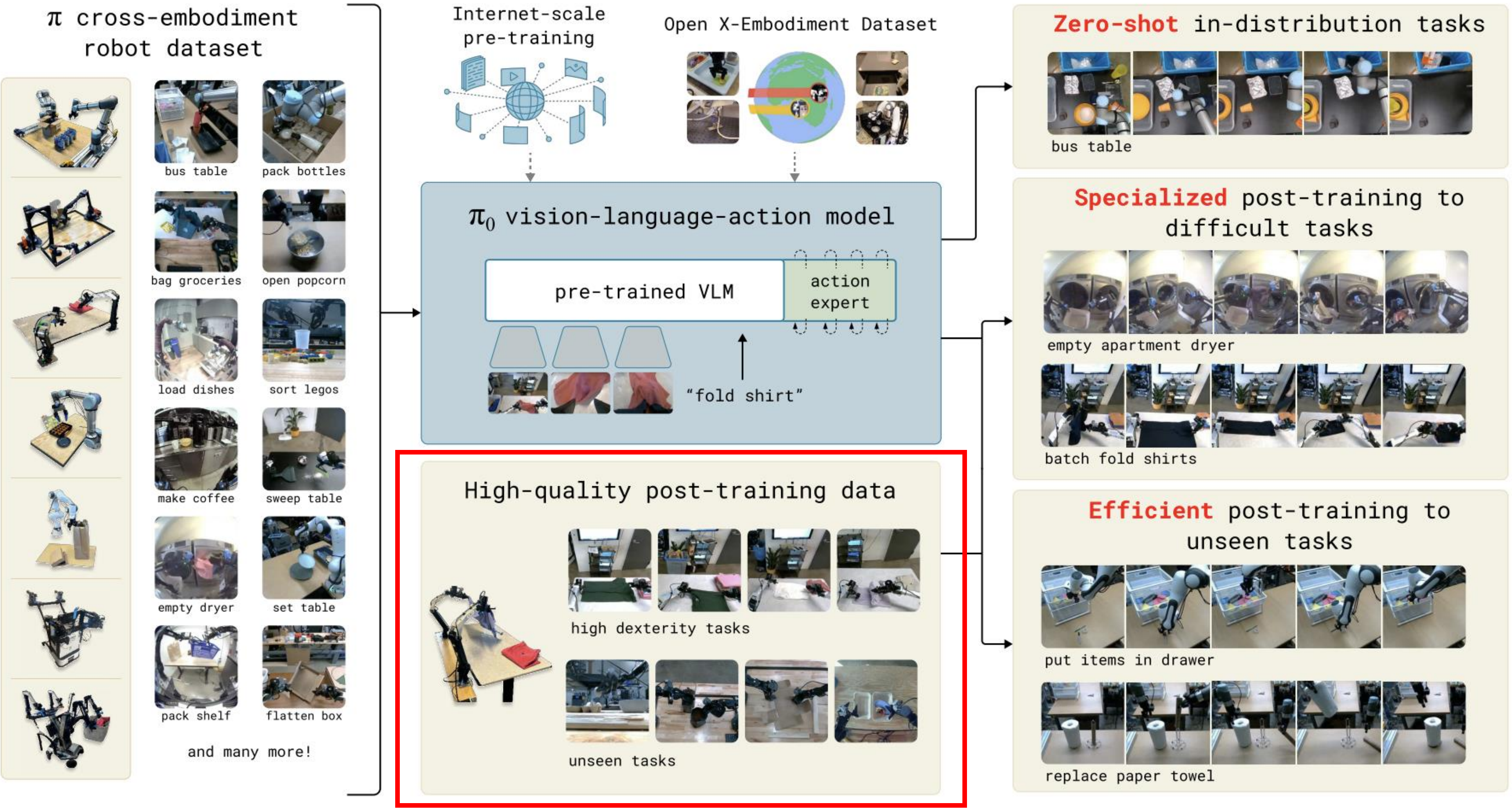


Cross-Embodiment Dataset

Pi-Zero by Physical Intelligence



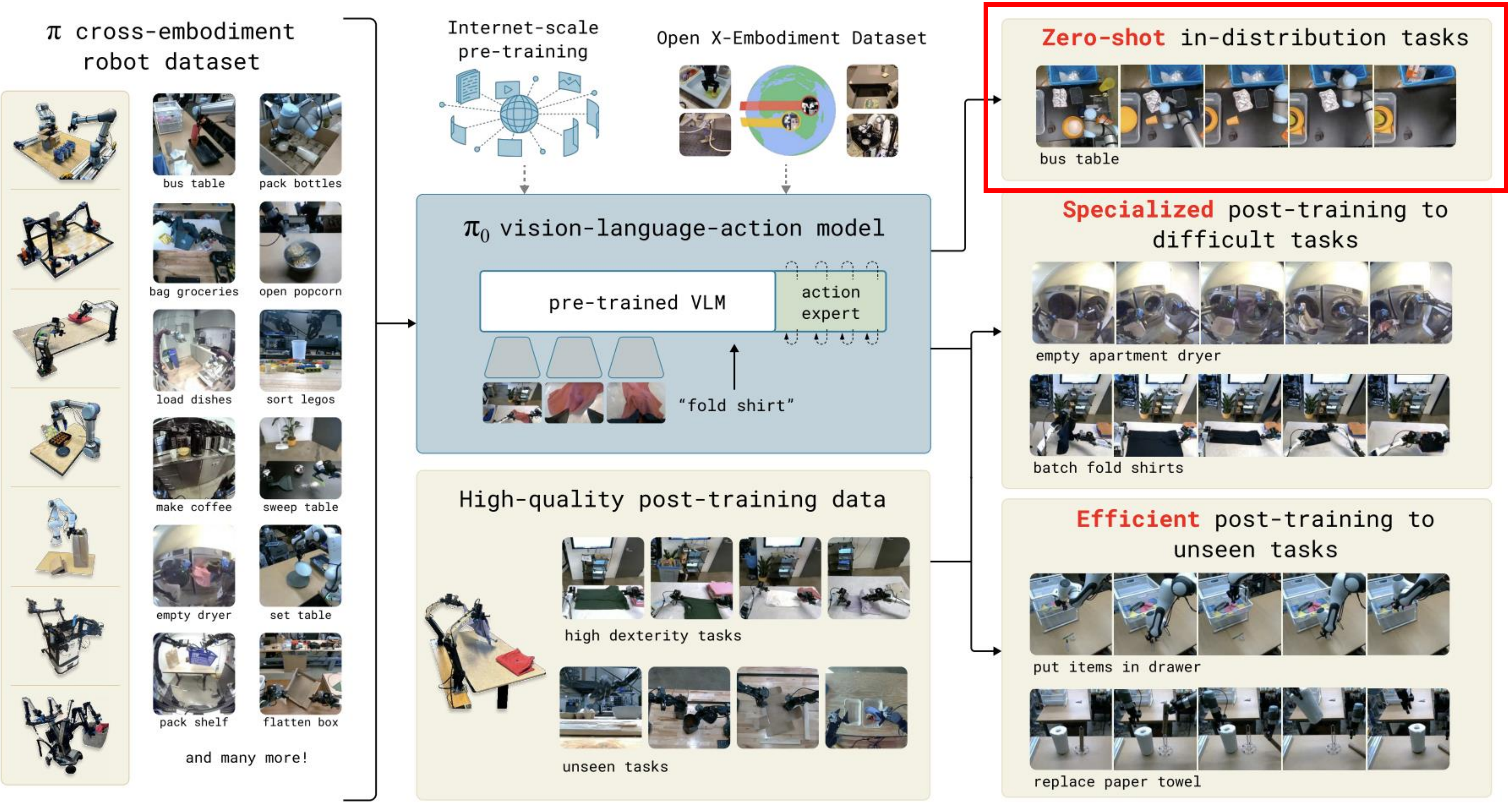
Pi-Zero by Physical Intelligence



Post-Training

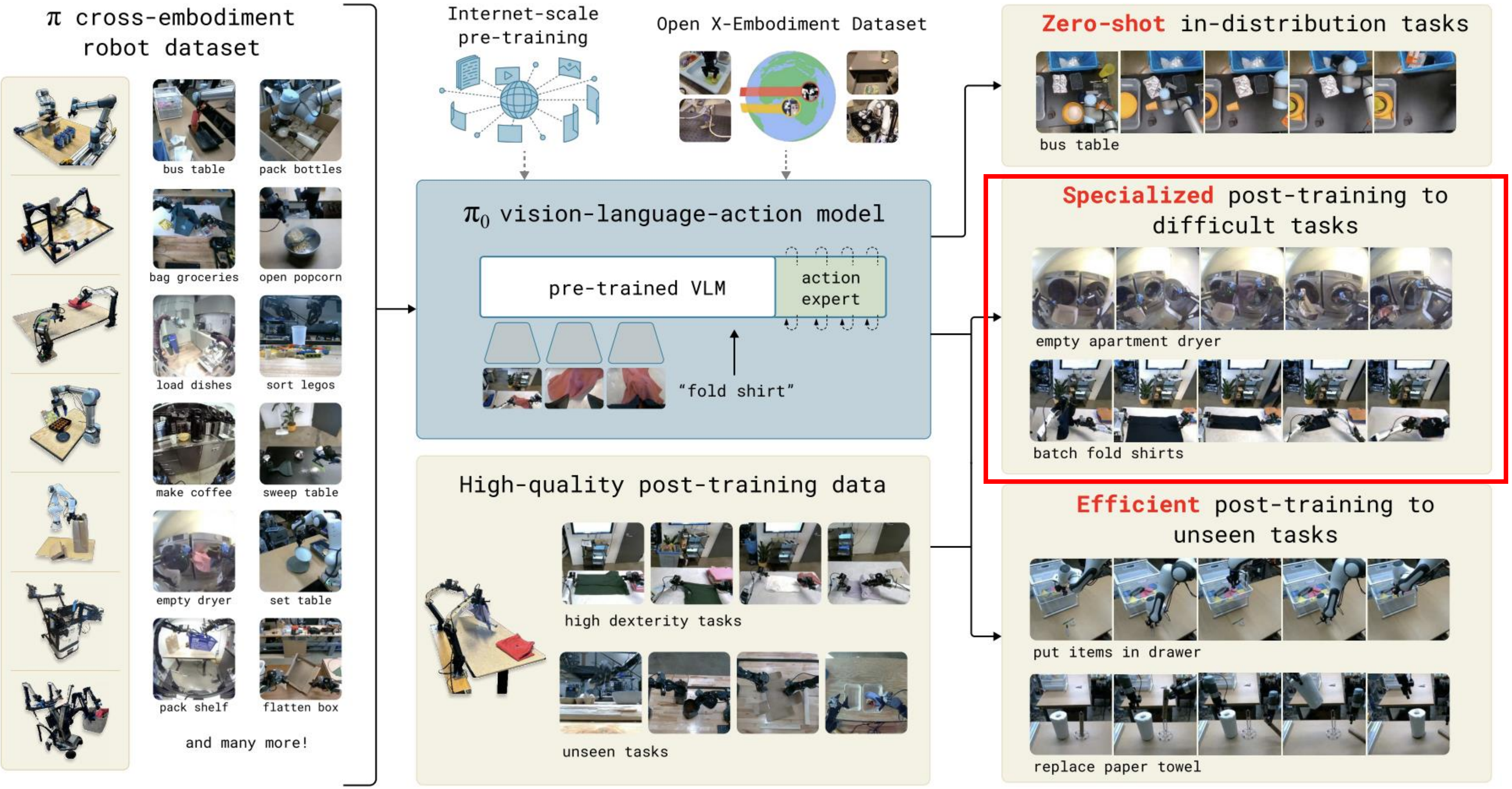
Pi-Zero by Physical Intelligence

Simple in-distribution tasks

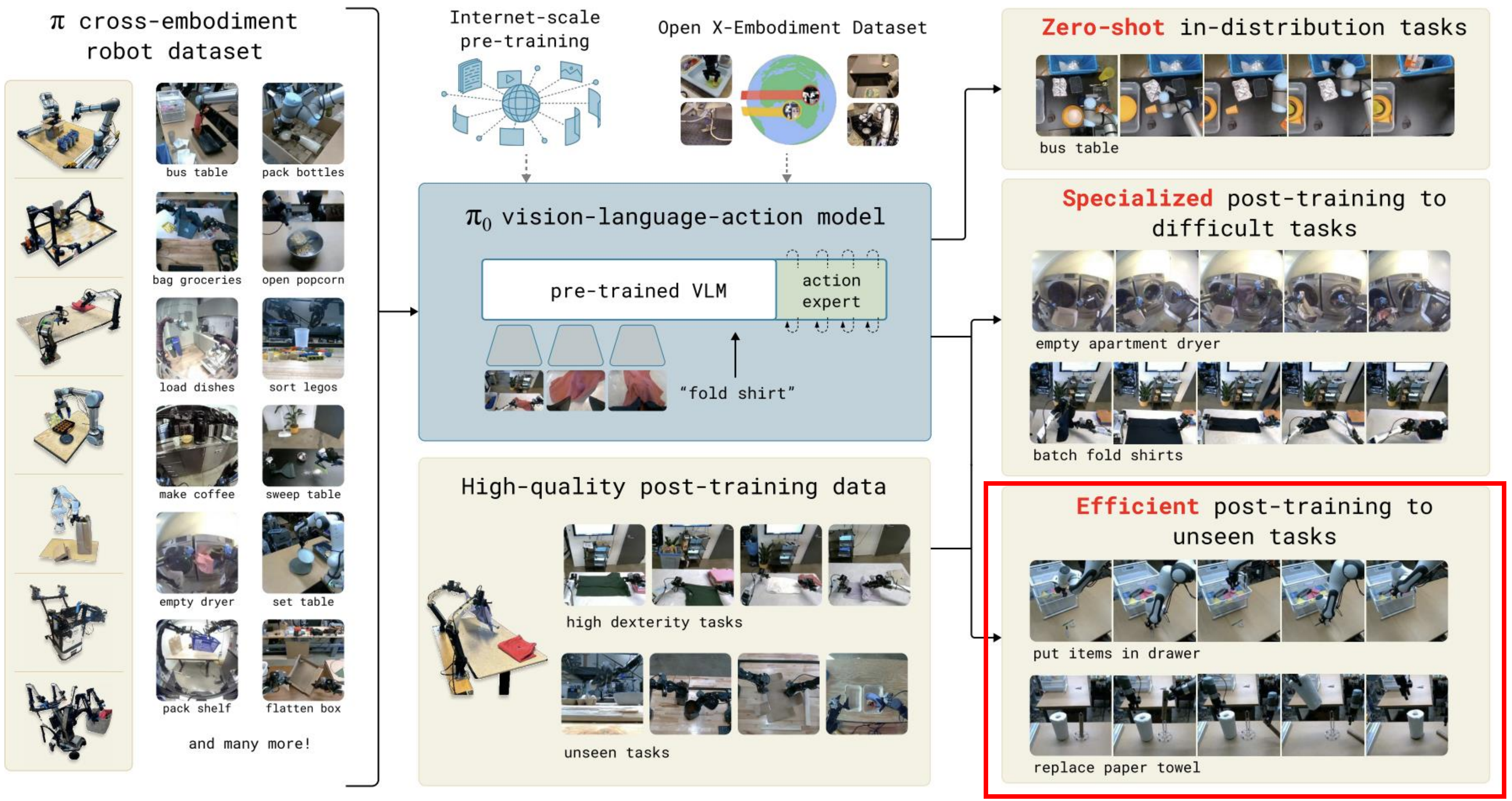


Pi-Zero by Physical Intelligence

Complicated in-distribution tasks

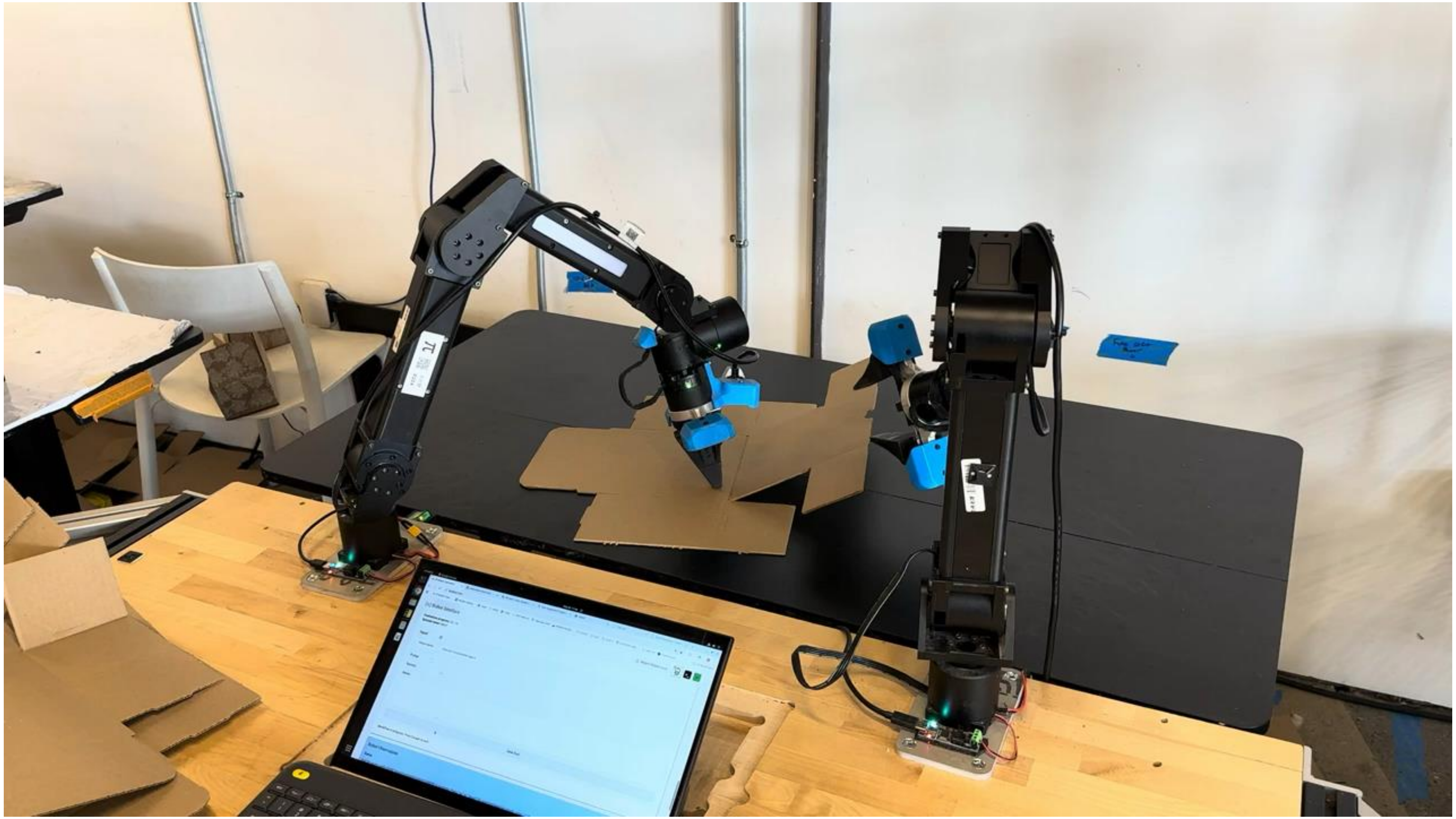


Pi-Zero by Physical Intelligence



Unseen tasks

Video I recorded yesterday at PI



Pi-Zero by Physical Intelligence

Physical Intelligence (π)

Open Sourcing π_0

Published	February 4, 2025
Email	research@physicalintelligence.company
Repo	Physical-Intelligence/openpi

README   Apache-2.0 license

openpi

openpi holds open-source models and packages for robotics, published by the Physical Intelligence team.

Currently, this repo contains two types of models:

- the π_0 model, a flow-based diffusion vision-language-action model (VLA)
- the π_0-FAST model, an autoregressive VLA, based on the FAST action tokenizer.

For both models, we provide *base model* checkpoints, pre-trained on 10k+ hours of robot data, and examples for using them out of the box or fine-tuning them to your own datasets.

This is an experiment: π_0 was developed for our own robots, which differ from the widely used platforms such as ALOHA and DROID, and though we are optimistic that researchers and practitioners will be able to run creative new experiments adapting π_0 to their own platforms, we do not expect every such attempt to be successful. All this is to say: π_0 may or may not work for you, but you are welcome to try it and see!

Overview

- Problem formulation
- Robot perception
- Reinforcement learning
- Model learning & model-based planning
- Imitation learning
- Robotic foundation models
- Remaining challenges

Evaluation of the Robot Learning Models

- Evaluation is primarily conducted in the real world
    - Real-world evaluation is costly and noisy
        - “We have large enough budget such that we can still make progress.”
    - Weak correlation between training loss and real-world success rate.
        - Training objectives vs task-specific metrics, training vs testing horizons

ALOHA 2

Evaluation of the Robot Learning Models

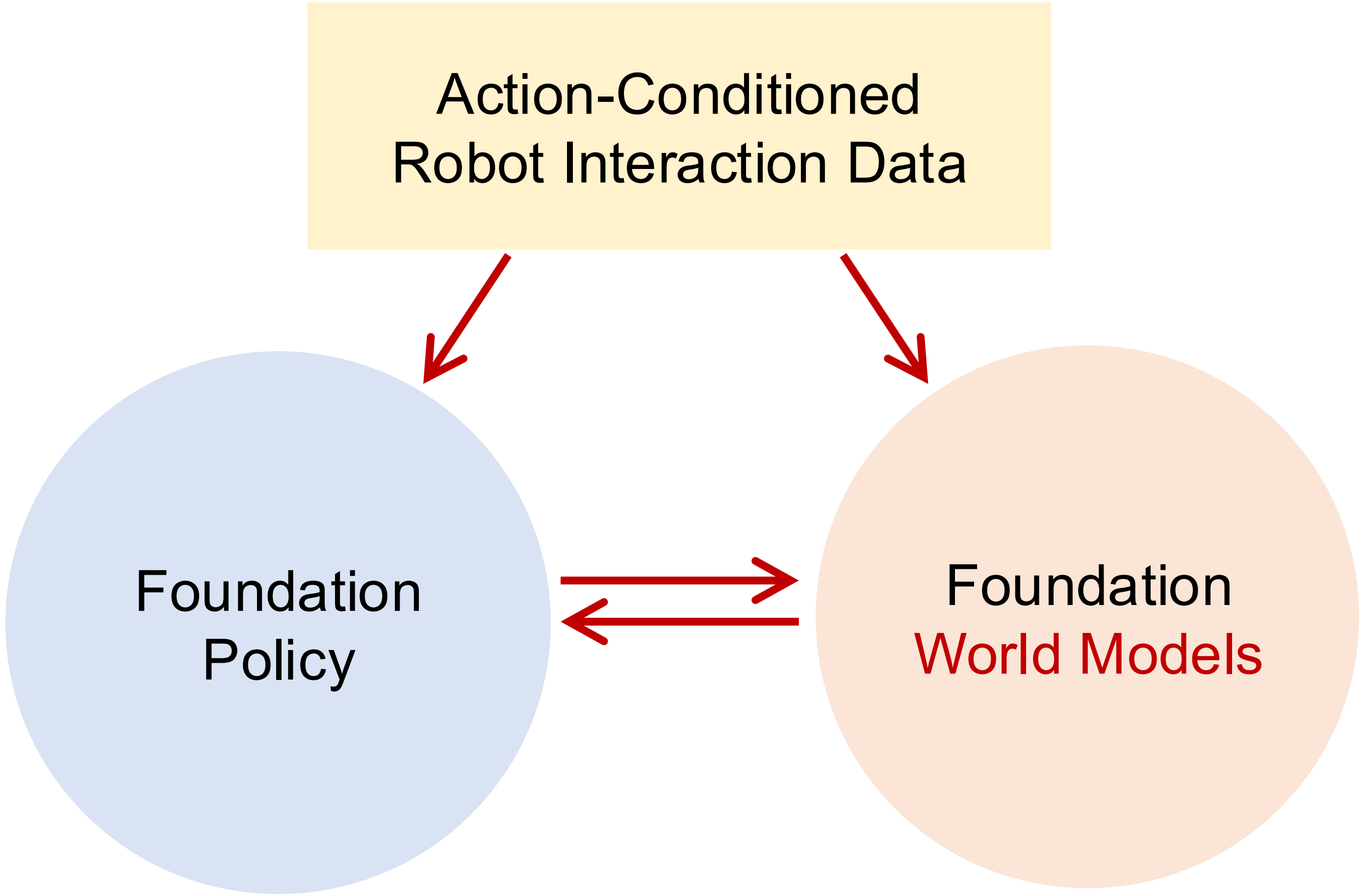
- What about evaluation in simulation?
  - Sim-to-real gap: rigid / deformable / cloth
  - Efficient asset generation
  - Digitalization of the real world
  - Procedural generation of realistic and diverse scenes
  - Correlation between sim and real

ImageNet in Embodied AI?

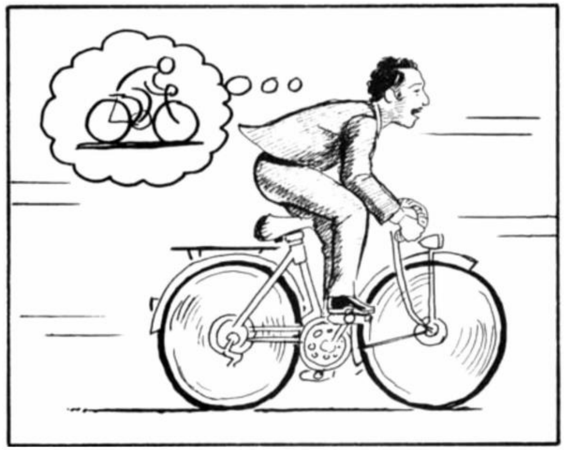
Habitat 3.0

Foundation Policy → Foundation World Models

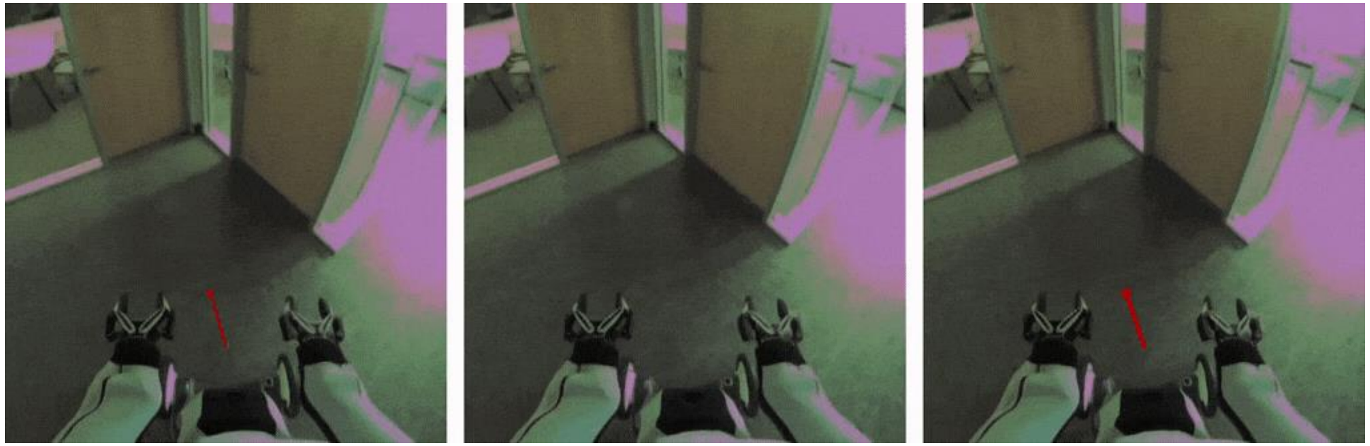
- My definition of world models: action-conditioned future prediction



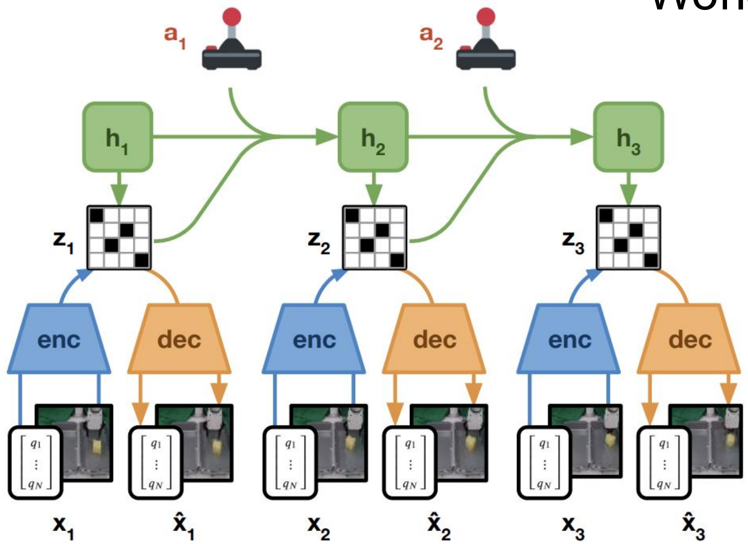
Foundation Policy → Foundation World Models



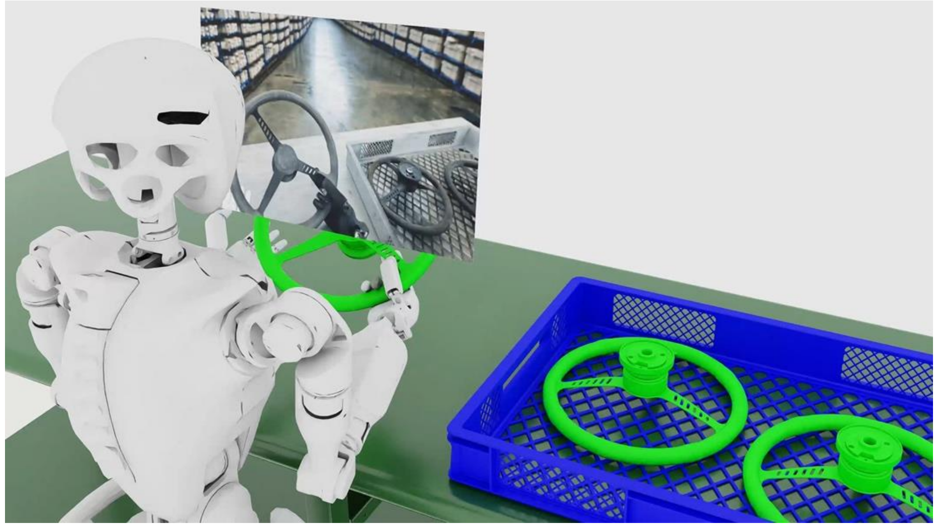
World Models



1X World Models



DayDreamer



Nvidia Cosmos - World Foundation Model

- 3D?
- Structural Prior?
- Learning + Physics?
- Corr. w/ Real World

Foundation Models for Embodied Agents

- Current foundation models are not tailored for embodied agents
  - LLM/VLM can fail in embodied-related tasks
  - Limited understanding of geometric / embodied / physical interactions
  - Reinforcement learning (RL) from human feedback → RL from Embodied Feedback

GPT

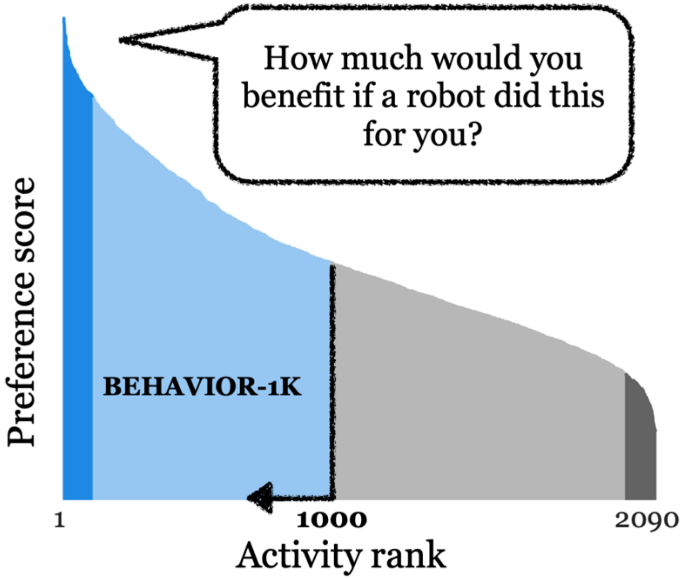
Segment Anything

DINOv2

Adaptation / Life-Long Learning

- Adapt to new scenarios
- Adapt to human preferences
- Self improve / life-long learning

Adapt to new scenarios



Adapt to human preferences

Improve through experience

Practical Considerations of Foundation Models

- Every robotics work is a system work
- System-level considerations: delays / computing / modules talking to each other

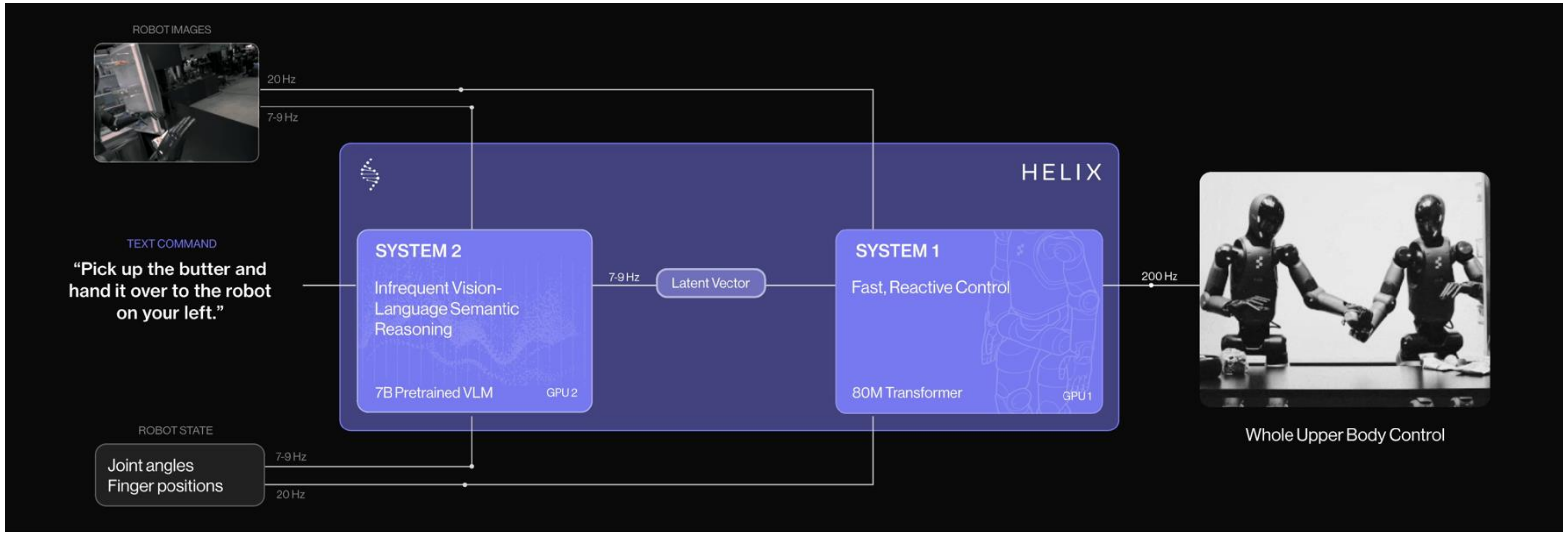
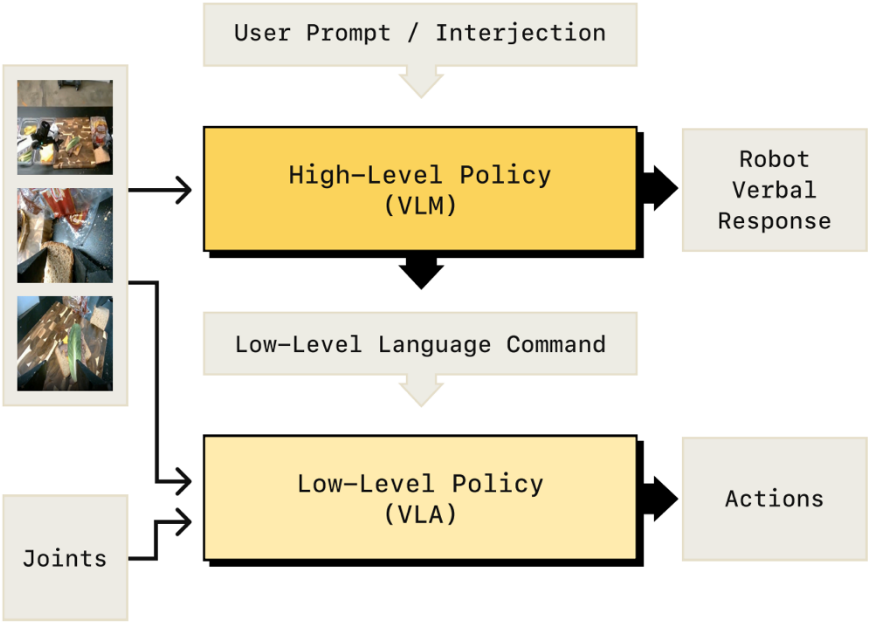


Figure AI: Helix



Physical Intelligence: Hi-Robot

Overview

- Problem formulation
- Robot perception
- Reinforcement learning
- Model learning & model-based planning
- Imitation learning
- Robotic foundation models
- Remaining challenges

Towards **foundational robotic models**








Images generated using AI

Next time: **Human-Centered AI**